



Alignement d'objets mécaniques complexes par vision monoculaire

Steve Bourgeois

► To cite this version:

Steve Bourgeois. Alignement d'objets mécaniques complexes par vision monoculaire. Optique / photonique. Université Blaise Pascal - Clermont-Ferrand II, 2006. Français. NNT : 2006CLF21697 . tel-00717568

HAL Id: tel-00717568

<https://theses.hal.science/tel-00717568>

Submitted on 13 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° d'ordre : D.U. 1797
EDSPIC : 358

UNIVERSITÉ BLAISE PASCAL - CLERMONT-FERRAND II

École Doctorale
Sciences Pour l'Ingénieur de Clermont-Ferrand

Thèse présentée par :
Steve BOURGEOIS

Formation Doctorale CSTI :
Composants et Systèmes pour le Traitement de l'Information

en vue de l'obtention du grade de

DOCTEUR D'UNIVERSITÉ

Spécialité : Vision pour la Robotique

**Alignement d'Objets Mécaniques Complexes par Vision
Monoculaire**

Soutenue publiquement le 28 Novembre 2006 devant le jury :

Mme Marie-Odile BERGER	Rapporteuse et Examinatrice
M. Michel DHOME	Directeur de thèse
M. Ernest HIRSCH	Rapporteur et Examineur
M. Frédéric JURIE	Invité
M. Vincent LEPETIT	Examineur
Mme Sylvie NAUDET COLLETTE	Examinatrice
M. Francis SCHMITT	Président du jury

Remerciements

Je remercie tout d’abord mes encadrants, c’est à dire Michel Dhome, mon directeur de thèse, et Sylvie Naudet-Collette. Ils m’ont proposé un sujet particulièrement intéressant et m’ont laissé une grande autonomie dans ma recherche tout en sachant me donner des conseils avisés aux bons moments.

Je remercie également Francis Schmitt, Ernest Hirsh, Marie Odile Berger et Vincent Lepetit pour avoir accepté de faire partie de mon jury, ainsi que Frédéric Jurie qui a accepté d’assister à ma soutenance de thèse.

Je remercie aussi la société ActiCM qui m’a fourni différentes séquences avec leurs vérités terrains.

Je tiens aussi à remercier de leur accueil les membres du laboratoire LCEI du CEA-LIST ainsi que ceux de l’équipe COMSEE du LASMEA. En particulier, je tiens à remercier Jean-Thierry Lapresté et Thierry Chateau pour leur remarques pertinentes sur mon manuscrit. Du côté du CEA-LIST, je tiens à remercier François Gaspard, Quoc-Cuong Pham, Patrick Sayd, Fabien Dekeyser et Sebastien Cornou, pour leurs nombreux conseils... et ceci même si certains ont essayé d’attenter à ma vie dans des rivières souterraines ou sur des glaciers ! ! ! Et je n’oublie pas non plus les thésard(e)s des différents labos : Laetitia Gond, Laetitia Leyrit, Etienne Mouragnon, Julien Bégard, Eric Royer, Julie Badrie, Hanna Martinsson, Stéphane Chevobbe, et bien d’autres encore...

A présent, passons à la partie un peu plus personnelle...

Comme on a rarement l’occasion de le graver dans le marbre (bien que je doute qu’on édite un jour mon manuscrit sur ce type de support, mais sait on jamais...) le nom des personnes auxquelles on tient, je profite de cette page pour citer un petit nombre d’amis qui ont particulièrement compté pour moi au cours de ces trois années de thèse.

Du côté de la région Parisienne, il y a tout d’abord Mr Philippe (hé, tu tiens

ta part de responsabilité dans cette thèse puisque c'est toi qui m'a refilé l'virus de l'informatique !), et l'un des deux seuls hommes-volatiles : Mr Thomas.

Du côté de Clermont, il y a tout d'abord la miss Hala, une poule naine insubmersible dans les eaux du Tarn. Mais il y a aussi m'zelle Titia (un conseil : arrête de gueuler à 4h du mat devant les fenêtres d'une bande de motards attardés ! Ou ça va finir par dégénérer !!!). Et pour finir, la star de la guitare connue sous le pseudonyme de "Roi du Banjo", c'est bien sûr Mr Gaetan !

Bien que ce soit généralement un passage inévitable dans les mémoires de thèse, je vais vous épargner l'énumération de la totalité des membres de ma famille (ce n'est pas "l'école des fans" après tout et Jacques Martin ne fait pas parti de mon jury). J vais donc me limiter à l'un de mes oncles : Guy Bourgeois, qui m'a vivement encouragé à entreprendre cette thèse. Pour autant, je n'oublie pas le reste de la famille...

Voilà voilà !!!! Désolé pour les autres, mais c'est pas le festival de Cannes, j'ai pas la soirée entière pour citer tout le monde...

Résumé

Ce mémoire présente un système de recalage visuel adapté aux pièces mécaniques. L'objectif de ces travaux est donc d'estimer, de manière automatique et précise, la position et l'orientation 3D d'un objet connu, à partir de son image vidéo. Les objets considérés dans le cadre de ces travaux possèdent la particularité d'être complexes, généralement métalliques et pas ou très peu texturés.

Pour répondre à cette problématique, nous proposons d'utiliser un modèle d'objet caractérisant à la fois l'apparence et la géométrie de l'objet. Un processus d'apprentissage permettant d'établir un tel modèle est proposé. Celui-ci exploite un nombre restreint de vues recalées avec l'objet, ainsi que le modèle 3D de ce dernier. Afin de pouvoir estimer la pose de l'objet, nous proposons un processus de mise en correspondance qui se déroule en deux étapes exploitant conjointement les informations d'apparence et de géométrie du modèle. La première étape utilise exclusivement l'apparence et repose sur un nouveau descripteur de contours. Nous montrons que ce seul processus est généralement insuffisant lorsqu'il est appliqué à des objets mécaniques complexes présentant des motifs répétitifs (série d'alésages,...). Ces derniers sont à l'origine d'un nombre de faux appariements suffisamment important pour qu'un calcul de pose ne puisse être envisagé en un temps raisonnable. Nous proposons donc une étape de filtrage géométrique des appariements exploitant la connaissance de la géométrie 3D de l'objet. Celle-ci repose sur des amers particuliers permettant d'obtenir une pose approximative de l'objet à partir d'un seul appariement. Une évaluation expérimentale indique que le taux de faux appariements, à l'issue du processus de filtrage, est suffisamment faible pour qu'un calcul de pose puisse être envisagé. Un processus d'estimation précise de la pose est alors proposé.

Enfin, une évaluation expérimentale du processus de recalage complet est présentée. Celle-ci a été menée sur des pièces mécaniques issues de l'industrie (portière, culasse de moteur,...). Les résultats obtenus soulignent la précision et la robustesse de l'approche proposée.

Abstract

this thesis presents a visual registration system suitable for mechanical objects. The aim of this work is to perform an accurate and automatic estimation of the 3D position and orientation of a known object, from a single camera image. The studied objects are complex, generally metallic and textureless.

To solve this problem, we introduce an object model which characterizes both the visual aspect and the 3D geometry of the object. This model is generated through a completely automatic learning process which exploits a small set of images of the object, registered with its CAD model.

To estimate the object position and orientation, we propose a two-step matching process which exploits both visual and geometric data of the object model. The first step is based on a new local contour descriptor which characterizes exclusively the visual aspect of the object. We show that such a process is generally insufficient when applied to complex mechanical objects. Most of the time, the presence of repetitive patterns (bore series, ...) generates such a large amount of false matches that a pose computation becomes computationally prohibitive. We suggest completing the matching process by a step of false matches filtering based on the knowledge of the 3D geometry. This process is based on a special feature which allows a pose computation from a single match. Experimental evaluations show that this filtering reduces the false match rate enough to envisage a pose computation. A pose estimation process is then explained.

Finally, an experimental evaluation of the complete visual registration process is exposed. This evaluation was lead on industrial mechanical objects (a car door, an automotive cylinder-head,...). The results underline the accuracy and the robustness of the proposed solution.

Table des matières

Introduction	1
1 Etat de l'Art et Contexte	5
1.1 Introduction	5
1.2 Approches 2D/2D	7
1.2.1 Choix des vues	8
1.2.2 Identification d'une vue à l'aide de primitives géométriques	11
1.2.3 Identification d'une vue à l'aide de primitives photomé-	
triques	14
1.2.4 Identification d'une vue à l'aide de descripteurs locaux . .	17
1.3 Approches 2D/3D	21
1.3.1 Choix des primitives du modèle	22
1.3.2 Mise en correspondance et estimation de la pose	23
1.4 Approches 3D/3D	26
1.5 Approches mixtes	28
1.6 Position des travaux présentés	29
2 Modèle combinant apparence et 3D	31
2.1 Les primitives visuelles	32
2.1.1 Région support : le voisinage d'un contour fermé	34
2.1.2 Caractérisation de la région support	38
2.2 Organisation des primitives visuelles en modèle géométrique 3D .	46
2.3 Extension du modèle : le motif 2D/3D	48
2.4 Conclusion	50
3 Mise en correspondance selon l'apparence	51
3.1 Processus de mise en correspondance	53
3.2 Modélisation de la variance de la signature	54
3.3 Gestion des motifs répétitifs	56

3.4	Évaluation	64
3.4.1	Protocole expérimental	64
3.4.2	Résultats	66
3.5	Conclusion	69
4	Filtrage des appariements et calcul de pose	71
4.1	Filtrage géométrique des appariements	72
4.1.1	Principe de la méthode	73
4.1.2	Estimation d'une pose à partir d'un seul appariement . . .	76
4.1.3	Critère de décision pour retenir les appariements cohé- rents avec une pose	83
4.1.4	Sélection des groupes d'appariements cohérents	88
4.2	Estimation Robuste de la pose	92
4.2.1	Calcul de la pose initiale	92
4.2.2	Raffinement de la pose	94
4.3	Conclusion	95
5	Évaluation du filtrage géométrique	97
5.1	Protocole expérimental	97
5.1.1	Les mesures	97
5.1.2	Les données	98
5.1.3	Les expériences	98
5.2	Pertinence du critère de cohérence	98
5.2.1	Le rappel	99
5.2.2	La précision	100
5.2.3	Le co-rappel	101
5.3	Performances du filtrage géométrique	105
5.3.1	Le rappel.	105
5.3.2	La précision en l'absence de pondération.	105
5.3.3	La précision avec pondération.	106
5.4	Conclusion	111
6	Évaluation du processus de recalage	113
6.1	Protocole expérimental	113
6.1.1	La base d'objets	113
6.1.2	Caractérisation de la qualité du recalage	118
6.1.3	Algorithmes évalués	121
6.2	Robustesse aux changements de points de vue	121
6.3	Performances avec un modèle léger	127

6.3.1	Échecs et réussites du recalage	127
6.3.2	Qualité visuelle du recalage.	128
6.3.3	Précision de la pose	133
6.3.4	Précision de la position 3D des points de surface visibles .	136
6.4	Tests dans un environnement non contrôlé	157
6.5	Conclusion	165
7	Conclusion et Perspectives	167
7.1	Un modèle caractérisant apparence et géométrie 3D	167
7.2	Un processus de mise en correspondance discriminant	168
7.3	Un recalage visuel précis et robuste	169
7.4	Perspectives	169
7.4.1	Amélioration du modèle	169
7.4.2	Amélioration du processus de mise en correspondance . .	170
7.4.3	Amélioration du calcul de pose	171
7.4.4	Extension du domaine d'application	171
A	Précision, Rappel et Courbe de Précision/Rappel	181
B	Distance des trièdres	185
C	Résultats du recalage avec un modèle léger	189

Table des figures

1.1	Exemple de graphe d'aspect	10
1.2	Exemple de sphère de vues	11
1.3	Exemple de représentation géométrique d'une vue	13
1.4	Comparaison entre l'utilisation de l'ACP et de la NMF pour la représentation de visages	17
1.5	Principe des approches 2D/3D	18
1.6	Principe des approches 2D/3D	22
1.7	Approche 3D/3D	27
1.8	Exemple de reconstruction 3D à l'aide de cylindres généralisés . .	27
2.1	Modèle combinant apparence et géométrie 3D	32
2.2	Instabilité de la texture des pièces mécaniques.	33
2.3	Régions extraites par un DoG-Affine sur une pièce mécanique . .	35
2.4	L'ellipse d'inertie d'un contour Γ	37
2.5	Calcul de la région support à partir d'un contour fermé.	38
2.6	Normalisation de la région support à une rotation près	40
2.7	Artefacts de segmentation lors de la normalisation d'une région support	41
2.8	Calcul du poids d'un point de contour	42
2.9	Pondération des points de contours	42
2.10	Calcul de la signature STHGI	43
2.11	Exemples de Région supports normalisées partageant la même si- gnature	44
2.12	Calcul de la signature <i>Shape Context</i>	45
2.13	Le modèle géométrique 3D	47
2.14	Principe du motif 2D/3D	49
3.1	Exemples de motifs répétitifs sur une culasse de moteur.	52
3.2	Mise en correspondance en deux passes	53
3.3	Approximation de la distribution des distances par une gaussienne	55

3.4	Exemple de distribution des distances à un descripteur	57
3.5	Exemple de distribution des distances pour une signature <i>Shape</i> <i>Context</i>	57
3.6	Première image d'apprentissage de la portière	58
3.7	Les régions supports normalisées extraites de la première image d'apprentissage de la culasse	59
3.8	Seconde image d'apprentissage de la portière	60
3.9	Les régions supports normalisées extraites de la seconde image d'apprentissage de la culasse	61
3.10	Troisième image d'apprentissage de la portière	62
3.11	Les régions supports normalisées extraites de la troisième image d'apprentissage de la culasse	63
3.12	Rappel du processus de mise en correspondance selon l'apparence	66
3.13	Précision du processus de mise en correspondance selon l'apparence	67
3.14	Courbe de précision/Rappel du processus de mise en correspondance selon l'apparence	69
4.1	Principe de la sélection des appariements selon un critère géométrique.	74
4.2	Les différentes étapes du filtrage géométrique	75
4.3	Lien entre les points 3D du motif 2D/3D et les points 2D de l'image.	77
4.4	Optimisation de la pose	78
4.5	Les erreurs mesurées pour caractériser la qualité d'une pose estimée à partir d'un seul appariement	81
4.6	Erreur sur la pose estimée à partir d'un seul appariement	82
4.7	Critère de cohérence d'un appariement 3D	86
4.8	Critères 2D et 3D pour l'identification d'appariements cohérents avec une pose	87
4.9	Phénomène de glissement du modèle CAO le long d'un axe de répétition des motifs répétitifs	93
5.1	Première expérience : Rappel(Θ_i) pour la portière	100
5.2	Première expérience : Moyenne et médiane de Précision(Θ_i) calculées sur les 50 vues de la portière.	101
5.3	Première expérience : Moyenne et médiane de CoRappel(Θ_i) calculées sur les 50 vues de la portière.	102
5.4	Résultats de la première expérience pour la Culasse.	103
5.5	Résultats de la première expérience pour l'étalon.	104

5.6	Seconde expérience : moyenne et médiane de $Rappel(\Xi_j)$ calculées sur les 50 vues de la portière.	106
5.7	Seconde expérience : moyenne et médiane de $Rappel(\Xi_j)$ calculées sur les 50 vues de la portière en l'absence de pondération des appariements	106
5.8	Influence de la pondération sur la précision	107
5.9	Seconde expérience : moyenne et médiane de $Precision(\Xi_j)$ calculées sur les 50 vues de la portière avec une prise en compte des poids de appariements.	107
5.10	Résultats de la seconde expérience pour l'étalon.	109
5.11	Résultats de la seconde expérience pour la culasse.	110
6.1	La culasse	115
6.2	La portière	116
6.3	L'Etalon	117
6.4	Représentation en fausse couleur de l'erreur de reprojection	119
6.5	Représentation en fausses couleurs de l'erreur de position	120
6.6	L'écart angulaire entre deux vues	122
6.7	Evaluation de la robustesse aux changements de points de vue pour la culasse	124
6.8	Evaluation de la robustesse aux changements de points de vue pour l'étalon	125
6.9	Evaluation de la robustesse aux changements de points de vue pour la portière	126
6.10	Configuration ambiguë pour l'étalon	128
6.11	Qualité visuelle du recalage	129
6.12	Exemples d'erreurs de reprojections.	130
6.13	Erreurs de translation	134
6.14	Erreurs de rotation	135
6.15	Erreurs de position 3D des points de surface visibles	137
6.16	Cas d'une erreur de reprojection faible malgré une erreur 3D non négligeable	138
6.17	Exemples de résultats obtenus sur la portière avec la méthode 2D/2D.	139
6.18	Exemples de résultats obtenus sur la portière avec la méthode 3D/2D.	140
6.19	Exemples de résultats obtenus sur la portière avec la méthode 3D/3D.	141
6.20	Erreur de reprojection du meilleur recalage de la portière	142

6.21	Erreur de position 3D des points de surface du meilleur recalage de la portière	143
6.22	Erreur de reprojection du plus mauvais recalage de la portière . . .	144
6.23	Erreur de position 3D des points de surface du plus mauvais recalage de la portière	145
6.24	Erreur de reprojection pour un recalage moyen de la portière . . .	146
6.25	Erreur de position 3D des points de surface pour un recalage moyen de la portière	147
6.26	Reprojection du modèle de la portière pour un recalage moyen selon la méthode 2D/2D.	148
6.27	Erreur de reprojection à la surface de la portière pour un recalage moyen selon la méthode 2D/2D	149
6.28	Erreur de position des points de surface de la portière pour un recalage moyen selon la méthode 2D/2D	150
6.29	Reprojection du modèle de la portière pour un recalage moyen selon la méthode 3D/2D.	151
6.30	Erreur de reprojection à la surface de la portière pour un recalage moyen selon la méthode 3D/2D	152
6.31	Erreur de position des points de surface de la portière pour un recalage moyen selon la méthode 3D/2D	153
6.32	Reprojection du modèle de la portière pour un recalage moyen selon la méthode 3D/3D.	154
6.33	Erreur de reprojection à la surface de la portière pour un recalage moyen selon la méthode 3D/3D	155
6.34	Erreur de position des points de surface de la portière pour un recalage moyen selon la méthode 3D/3D	156
6.35	Base d'images pour l'évaluation en environnement non contrôlé .	158
6.36	Recalages selon l'approche 2D/2D dans un environnement non contrôlé	159
6.37	Recalages selon l'approche 3D/2D dans un environnement non contrôlé	160
6.38	Recalages selon l'approche 3D/3D dans un environnement non contrôlé	161
6.39	Exemple de recalage de la culasse selon l'approche 2D/2D dans un environnement non contrôlé.	162
6.40	Exemple de recalage de la culasse selon l'approche 3D/2D dans un environnement non contrôlé.	163
6.41	Exemple de recalage de la culasse selon l'approche 3D/3D dans un environnement non contrôlé.	164

A.1	L'espace des appariements	181
A.2	Calcul de la précision	182
A.3	Variation de la précision en fonction de la taille K de l'ensemble d'appariements établis dans le cas d'une sélection idéale.	182
A.4	Calcul du Rappel	183
A.5	Variation du rappel en fonction de la taille K de l'ensemble des appariements établis dans le cas d'une sélection idéale.	183
A.6	Courbe de Précision/Rappel dans le cas d'une sélection idéale. . .	184
B.1	Principe de l'erreur des trièdres.	185
B.2	Pondération des composantes rotationnelles et translationnelles dans la distance du trièdre.	186
C.1	Echelle de l'erreur de reprojection	189
C.2	Echelle de l'erreur de position 3D	189
C.3	Résultats de l'évaluation sur l'étalon	190
C.4	Exemples de résultats pour l'étalon avec la méthode 2D/2D. . . .	191
C.5	Exemples de résultats pour l'étalon avec la méthode 3D/2D. . . .	192
C.6	Exemples de résultats pour l'étalon avec la méthode 3D/3D. . . .	193
C.7	Erreur de reprojection du meilleur recalage de l'étalon	194
C.8	Erreur de position 3D des points de surface du meilleur recalage de l'étalon	195
C.9	Erreur de reprojection du plus mauvais recalage de l'étalon	196
C.10	Erreur de position 3D des points de surface du plus mauvais reca- lage de l'étalon	197
C.11	Erreur de reprojection pour un recalage moyen de l'étalon	198
C.12	Erreur de position 3D des points de surface pour un recalage mo- yen de l'étalon	199
C.13	Reprojection du modèle de l'étalon pour un recalage moyen selon la méthode 2D/2D.	200
C.14	Erreur de reprojection à la surface de l'étalon pour un recalage moyen selon la méthode 2D/2D	201
C.15	Erreur de position des points de surface de l'étalon pour un reca- lage moyen selon la méthode 2D/2D.	202
C.16	Reprojection du modèle de l'étalon pour un recalage moyen selon la méthode 3D/2D.	203
C.17	Erreur de reprojection à la surface de l'étalon pour un recalage moyen selon la méthode 3D/2D.	204

C.18	Erreur de position des points de surface de l'étalon pour un recalage moyen selon la méthode 3D/2D	205
C.19	Reprojection du modèle de l'étalon pour un recalage moyen selon la méthode 3D/3D.	206
C.20	Erreur de reprojection à la surface de l'étalon pour un recalage moyen selon la méthode 3D/3D	207
C.21	Erreur de position des points de surface de l'étalon pour un recalage moyen selon la méthode 3D/3D	208
C.22	Résultats de l'évaluation sur la culasse	210
C.23	Exemples de résultats pour la culasse avec la méthode 2D/2D. . .	211
C.24	Exemples de résultats pour la culasse avec la méthode 3D/2D. . .	212
C.25	Exemples de résultats pour la culasse avec la méthode 3D/3D. . .	213
C.26	Erreur de reprojection du meilleur recalage de la culasse	214
C.27	Erreur de position 3D des points de surface du meilleur recalage de la culasse	215
C.28	Erreur de reprojection du plus mauvais recalage de la culasse . . .	216
C.29	Erreur de position 3D des points de surface du plus mauvais recalage de la culasse	217
C.30	Erreur de reprojection pour un recalage moyen de la culasse . . .	218
C.31	Erreur de position 3D des points de surface pour un recalage moyen de la culasse	219
C.32	Reprojection du modèle de la culasse pour un recalage moyen selon la méthode 2D/2D.	220
C.33	Erreur de reprojection à la surface de la culasse pour un recalage moyen selon la méthode 3D/3D	221
C.34	Erreur de position des points de surface de la culasse pour un recalage moyen selon la méthode 2D/2D.	222
C.35	Reprojection du modèle de la culasse pour un recalage moyen selon la méthode 3D/2D.	223
C.36	Erreur de reprojection à la surface de la culasse pour un recalage moyen selon la méthode 3D/2D	224
C.37	Erreur de position des points de surface de la culasse pour un recalage moyen selon la méthode 3D/2D	225
C.38	Reprojection du modèle de la culasse pour un recalage moyen selon la méthode 3D/3D.	226
C.39	Erreur de reprojection à la surface de la culasse pour un recalage moyen selon la méthode 3D/3D	227
C.40	Erreur de position des points de surface de la culasse pour un recalage moyen selon la méthode 3D/3D.	228

Introduction

Au cours des années 80, l'industrie a connu une évolution marquante avec l'essor de la Conception et Fabrication Assistée par Ordinateur, ou CFAO. Ces nouvelles méthodes de conception et de production industrielles reposent sur la notion de *maquette numérique*, aussi nommée *modèle CAO*. Cette maquette numérique correspond à un prototype virtuel d'un objet. Celle-ci est obtenue à l'aide d'outils de CAO (Conception Assistée par Ordinateur), ces derniers permettant aussi de visualiser cette maquette et d'analyser son comportement sous diverses contraintes au travers de simulations numériques. Lors de l'étape de conception, la maquette numérique permet donc d'élaborer rapidement un prototype de l'objet, de le tester et de le modifier en fonction des résultats des tests sans avoir à produire un prototype réel. Lors de l'étape de fabrication, la maquette numérique est exploitée par des logiciels de FAO (Fabrication Assistée par Ordinateur) pour commander de manière automatique les machines outils. La maquette numérique tient donc une place clé depuis la conception de l'objet jusqu'à la réalisation de celui-ci. Cependant, son rôle en dehors de ce cadre est quasi-inexistant. Pourtant, de nombreuses applications peuvent tirer parti de la connaissance contenue dans le modèle CAO. Un premier domaine d'application correspond à celui du contrôle dimensionnel avec, par exemple, le contrôle de conformité ou le contrôle d'usure d'une pièce. Un autre domaine correspond à celui de la robotique, avec des applications comme la préhension d'une pièce par un bras robotisé. Enfin l'assistance, avec des applications comme l'aide à la maintenance par des techniques de réalité augmentée, peut avantageusement exploiter ce type de données.

Un frein au développement de ces applications vient des techniques actuellement employées pour exploiter le modèle CAO. Ainsi, le domaine du contrôle dimensionnel emploie principalement des Machines de Mesures Tridimensionnelles (ou MMT) mécaniques. Il s'agit de techniques de mesure avec contact coûteuses. De plus, ces techniques sont lentes, ce qui empêche de contrôler chaque pièce produite. Les contrôles sont alors effectués par lot. Le contrôle est réalisé sur la dernière pièce de chaque lot, et l'acceptation du lot complet dépend du résultat de

ce contrôle. D'autre part, les mesures ne peuvent être réalisées que si la pièce est placée sur la MMT selon une pose prédéterminée.

Dans le domaine de la réalité augmentée, de nombreuses solutions ont été développées à l'aide de capteurs de position et d'orientation magnétiques, acoustiques ou inertiels. Ces techniques sont généralement peu adaptées au milieu industriel puisque celles-ci nécessitent l'instrumentation de l'environnement, fonctionnent dans un volume restreint et pour un environnement contraint (e.g. les capteurs magnétiques nécessitent un environnement dépourvu d'objets métalliques et de toute source magnétique pour pouvoir assurer la précision des mesures).

Le développement des techniques de vision par ordinateur a permis de proposer de nouvelles solutions pour ces applications, celles-ci étant généralement plus souples d'utilisation, plus rapide et moins coûteuses.

Cependant, les solutions industrielles conservent la plupart du temps une contrainte : l'instrumentation de l'environnement par des marqueurs visuels dans le but de localiser la pièce étudiée. C'est pourquoi le problème de la localisation d'un objet, c'est-à-dire l'estimation de la position et de l'orientation de l'objet par rapport à la caméra, à partir de primitives visuelles naturelles a été particulièrement étudié ces dernières années. De nombreuses solutions ont été développées pour les objets plans, et appliquées, par exemple, au tri de pièces mécaniques planes par vision. S'agissant de pièces non-planes, des solutions adaptées à des pièces simples (généralement polyédriques) non-texturées, ou aux pièces complexes texturées ont été proposées. Cependant, les pièces mécaniques correspondent généralement à des pièces complexes non-texturées. Les solutions actuelles sont donc peu adaptées à cette problématique. Les travaux présentés dans ce mémoire proposent une solution de *recalage visuel* (aussi appelée *alignement*) adaptée aux pièces mécaniques complexes.

Objectif et méthode utilisée

L'objectif de ces travaux est de proposer un système de recalage visuel adapté aux pièces mécaniques, celles-ci étant généralement non planes, complexes, peu texturées et métalliques. Il s'agit aussi de définir une solution adaptée à un environnement non contrôlé. Le système doit être robuste aux variations d'illumination ainsi qu'à la présence d'objets parasites, ces derniers pouvant entraîner des occultations partielles de l'objet. De plus, la méthode proposée doit permettre d'obtenir un recalage visuel précis.

La solution employée consiste à identifier certaines manifestations visuelles de

l'objet à l'image, à déterminer leur position dans un modèle 3D de l'objet, pour enfin en déduire la position et l'orientation 3D de l'objet par rapport à la caméra.

Contributions

Les travaux présentés dans ce mémoire s'articulent autour de trois principales étapes : l'étape de création du modèle de l'objet, celle de mise en correspondance des primitives de l'image avec celles du modèle de l'objet, et enfin le calcul de la pose. Les principales contributions de ce travail concernent :

- La définition d'un modèle, caractérisant à la fois l'apparence et la géométrie de l'objet, adapté aux pièces mécaniques pas ou peu texturées. Ce modèle est établi à partir du modèle CAO de l'objet et d'un nombre restreint de vues de l'objet recalées avec le modèle CAO.
- Un processus de mise en correspondance entre les primitives visuelles d'une image et celles du modèle d'un objet. Ce processus doit établir un ensemble d'appariements comportant une faible fraction d'appariements erronés, ceci en dépit de la présence de nombreux motifs répétitifs.
- Une évaluation du système sur des scènes réelles mettant en oeuvre des pièces mécaniques complexes issues de l'industrie automobile.

Les travaux réalisés au cours de cette thèse ont donné lieu à plusieurs publications [9, 10, 11, 12].

Structure du document

Le premier chapitre de ce manuscrit est consacré à un état de l'art des méthodes de recalage visuel. Nous tâcherons également de situer la méthode proposée par rapport aux solutions existantes. Le second chapitre détaille le modèle utilisé pour représenter à la fois l'apparence et la géométrie de l'objet.

Dans le chapitre 3, un processus de mise en correspondance exploitant l'apparence de l'objet, et permettant d'établir une correspondance entre les primitives du modèle et celles issues d'une image, est présenté.

Le chapitre 4 explique comment l'information géométrique contenue dans le modèle de l'objet est employée pour filtrer les appariements et calculer la pose de l'objet. Une évaluation du processus de filtrage est présentée dans le chapitre 5 alors que le processus de recalage visuel complet est évalué dans le chapitre 6.

Chapitre 1

Etat de l'Art et Contexte

1.1 Introduction

Le **recalage visuel** (aussi appelé alignement par vision) consiste à estimer la position et l'orientation d'un objet tridimensionnel à partir d'une ou plusieurs images de celui-ci. Ce chapitre présente un état de l'art des méthodes de recalage visuel mono-image exploitant un modèle, c'est à dire des méthodes estimant la pose de l'objet à partir d'une seule image et à l'aide de connaissances a priori sur l'objet. Ces images sont issues du capteur plan d'une caméra dont on suppose qu'elle respecte le modèle de projection perspective.

La projection d'un objet sur une caméra engendre des variations de luminosité et de couleur dans l'image. Ces variations propres à la projection d'un objet sont désignées sous le terme de **manifestations visuelles** de l'objet. L'ensemble de celles-ci forme l'image de l'objet, ou **aspect** de l'objet à l'image. Une scène ne comportant généralement pas qu'un seul objet, le recalage d'un objet particulier consiste, dans un premier temps, à extraire et identifier les manifestations visuelles de celui-ci. Dans un second temps, il s'agit d'inférer la pose de cet objet à partir de ces manifestations visuelles et à l'aide d'un ensemble de connaissances a priori sur l'objet : le **modèle de l'objet**.

Une première difficulté de cette tâche provient des variations d'aspects que peut présenter un objet dans différentes images. Celles-ci dépendent :

- du point de vue : l'objet étant tridimensionnel, les parties visibles de celui-ci, ainsi que la manière dont celles-ci se projettent à l'image, dépendent du point de vue depuis lequel il est observé.
- des paramètres intrinsèques de la caméra : il s'agit des paramètres caractérisant la caméra (distance focale, point principal, ...). Le processus de pro-

jection dépend des valeurs de ceux-ci. Cependant, on considérera dans le reste du chapitre que la caméra est calibrée, c'est à dire que ses paramètres intrinsèques sont connus.

- des conditions d'illumination : les zones d'ombre et de lumière sur l'objet dépendent de l'éclairage de celui-ci.
- de la couleur et de la texture : plusieurs instances d'un même objet peuvent présenter des textures et des couleurs différentes (eg. un même modèle de voiture peut être décliné sous différentes couleurs).
- de la présence d'occultations : l'objet n'étant pas nécessairement isolé, les autres éléments de la scène peuvent masquer certaines parties de l'objet.

D'autre part, l'identification de manifestations visuelles associées à l'objet nécessite de séparer celles-ci des manifestations des autres éléments de la scène. Grimson [30] présente d'ailleurs cette étape comme le point le plus délicat du processus de recalage.

Les premières solutions ont été proposées avec les travaux de Roberts [70], ceux-ci traitant du problème du recalage visuel d'objets polyédriques. Depuis, de nombreuses approches ont été développées. Généralement, ces méthodes se différencient les unes des autres selon :

- la classe d'objets considérée : il n'existe pas de solution universelle, chaque solution est adaptée à une classe d'objet plus ou moins large (eg. objets polyédriques, objets texturés, ...).
- la nature de la connaissance a priori utilisée : certaines solutions vont exploiter la connaissance de la géométrie de l'objet, d'autres la connaissance de l'apparence de l'objet et d'autres encore vont s'aider d'informations sur le contexte¹ de la scène (eg. l'environnement présent dans l'image) [79].
- la robustesse aux changements environnementaux : certaines solutions nécessitent le contrôle de l'environnement (l'objet doit être isolé, l'éclairage contrôlé, ...) alors que d'autres n'imposent aucune contrainte (la scène peut comporter de nombreux objets parasites, l'objet peut être partiellement occulté, ...).
- l'efficacité : le coût en espace mémoire et en temps de calcul peut limiter le champ d'application de la solution (est-ce une solution «temps réel ?», le système est-il capable de recalculer plus d'un objet ?).
- la qualité du recalage : bien qu'il s'agisse d'une notion mal définie, on peut néanmoins différencier les approches estimant un recalage grossier (pose approximative de l'objet) des solutions estimant un recalage fin. Ces solu-

¹Les méthodes exploitant le contexte environnemental ne sont pas considérées dans cet état de l'art.

tions n'ont pas les mêmes champs d'application (un recalage approximatif n'est pas envisageable dans le cadre de métrologie par vision).

Dans la suite de ce chapitre, les méthodes sont organisées selon la manière d'identifier les primitives visuelles. Elles se répartissent alors selon quatre catégories :

- les méthodes 2D/2D : les primitives visuelles (2D) issues de l'image sont appariées aux primitives 2D du modèle de l'objet ;
- les méthodes 2D/3D : les primitives visuelles (2D) issues de l'image sont appariées à des primitives 3D du modèle de l'objet ;
- les méthodes 3D/3D : les primitives visuelles (2D) issues de l'image sont utilisées pour inférer la géométrie des primitives 3D qui les ont générées ; ces dernières étant ensuite appariées aux primitives 3D du modèle de l'objet ;
- les méthodes mixtes : il s'agit d'approches combinant au moins deux des trois premières approches.

1.2 Approches 2D/2D

Les approches 2D/2D visent à transformer le problème du recalage d'un objet 3D en un problème de recalage d'un objet 2D. Pour cela, ces approches utilisent la connaissance a priori des différents aspects pouvant être pris par l'objet à l'image : il s'agit d'un *modèle d'apparence*. Puisque l'aspect d'un objet à l'image est une description bidimensionnelle de l'objet, on peut considérer que ces approches représentent l'objet, non pas comme une entité 3D, mais comme une entité 2D dont l'aspect peut varier. Le recalage consiste alors à identifier l'aspect pris par cette entité 2D ainsi que sa position et son orientation dans l'image.

Pour pouvoir néanmoins déterminer la pose 3D d'un objet, les approches 2D/2D exploitent le fait que l'aspect de celui-ci est généralement sensible aux variations de points de vue. L'élaboration du modèle d'apparence consiste alors à déterminer l'ensemble des aspects pouvant être pris par l'objet et à attribuer à chacun d'eux le point de vue depuis lequel cet aspect est observé. L'identification de l'aspect pris par l'objet suffit alors à déterminer sa pose.

Puisque la détermination de la pose 3D de l'objet est implicite, le processus de recalage se résume donc à la seule étape d'extraction des manifestations visuelles correspondant à l'objet et à l'identification, à partir de celles-ci, de l'aspect pris par l'objet. Le modèle d'apparence et le processus d'élaboration de celui-ci sont

donc les éléments fondamentaux de cette approche.

Le modèle d'apparence d'un objet correspond généralement à une collection de vues de l'objet, chaque vue correspondant à une observation de l'objet dans une pose et/ou des conditions d'illuminations différentes. Pour que le modèle d'apparence puisse être représenté en mémoire, il est constitué d'un nombre fini de vues : les *vues caractéristiques*. Pour représenter l'aspect de l'objet dans une vue caractéristique, des manifestations visuelles de types particuliers (e.g. les contours, les points d'intérêt, ...), appelées **primitives visuelles**, sont extraites de chacune des vues. L'ensemble de ces primitives visuelles ainsi que leur configuration spatiale caractérisent alors l'aspect de l'objet.

L'ensemble des vues caractéristiques et leur description forment ainsi le modèle de l'objet : on parle alors de modèle *multi-vue*.

Les deux étapes de l'élaboration d'un modèle sont donc :

- le choix des vues caractéristiques,
- la caractérisation des vues à l'aide de primitives visuelles.

1.2.1 Choix des vues

Pour représenter la totalité des aspects pouvant être pris par un objet, il suffit de générer une vue caractéristique pour chaque point de vue depuis lequel l'objet est observable. Cet ensemble de vues est désigné sous le terme de *sphère de vues complète*.

Dans la pratique, la sphère de vues complète ne peut être employée puisque celle-ci contient une infinité de vues. Le modèle correspond donc à un sous-ensemble fini de la sphère de vues complète. Pour que le modèle de l'objet soit complet, il faut que ce passage à un sous-ensemble fini de vues n'entraîne pas une perte d'information dans la description de l'objet. Le choix des vues formant ce sous-ensemble est donc primordial.

D'autre part, à l'issue du recalage, la pose estimée correspond généralement à celle de la vue du modèle dont l'aspect est le plus proche de l'image observée. Il s'agit donc d'un recalage approximatif dont la précision dépend du nombre de vues que comporte le modèle. Il existe donc un compromis entre la taille du modèle et la précision du recalage. Dans certains cas, il est d'ailleurs possible d'estimer la taille du modèle selon la précision de recalage souhaitée [13].

Dans la pratique, les approches 2D/2D sont généralement employées pour estimer une pose grossière de l'objet. Il s'agit alors d'élaborer un modèle permettant de représenter l'ensemble des aspects de l'objet avec un minimum de vues. Les

méthodes permettant de déterminer ces vues se répartissent alors en deux catégories :

- les méthodes analytiques : celles-ci exploitent la connaissance du modèle CAO de l'objet afin de pouvoir déterminer précisément les vues caractéristiques pertinentes et les générer synthétiquement,
- les méthodes empiriques : celles-ci consistent généralement à établir un modèle à partir d'un grand nombre de vues puis à réduire la taille du modèle en fusionnant les vues partageant une apparence similaire ².

Méthodes analytiques

Afin de choisir les vues caractéristiques, les méthodes analytiques procèdent à un découpage en régions de la sphère de vues complète. Ce découpage est réalisé de manière à obtenir un nombre minimal de régions, chacune de ces régions devant contenir des points de vue pour lesquels l'aspect de l'objet est similaire³. Ce découpage en région de la sphère de vues peut être considéré comme un découpage en classes d'équivalences, deux points de vue étant considérés équivalents si les deux conditions suivantes sont respectées :

- les deux points de vue ont le même aspect,
- il existe un chemin sur la sphère de vues complète qui relie les deux points de vue et qui passe uniquement par des points de vue partageant le même aspect.

Puisque tous les points de vue d'une même région partagent un aspect similaire, chaque région peut être représentée par l'aspect d'une seule de ces vues sans entraîner une perte d'information. Il suffit alors de relier par une arête les vues issues de régions voisines pour obtenir une représentation de l'objet sous forme de graphe : *le graphe d'aspect*. La figure 1.1 représente le graphe d'aspect d'un cube.

On notera cependant que cette description n'est pas complète. En effet, les points de vue situés aux frontières des régions correspondent aux *points de vue accidentels*. Il s'agit de points de vue pour lesquels un changement infinitésimal de la pose entraîne un changement notable de l'apparence de l'objet (eg. un changement de topologie au niveau des primitives). Puisqu'un point de vue accidentel n'appartient à aucune région, l'aspect pris par l'objet pour ce point de vue n'est pas représenté dans le graphe d'aspect de l'objet. Néanmoins, la probabilité

²La notion de similarité d'aspect entre deux vues dépend de la manière dont est caractérisée une vue (cf. sections 1.2.2 et suivantes). Cependant, il s'agit généralement d'estimer si les primitives des deux images peuvent être mises en correspondance avec une précision élevée.

³Généralement, dans le cas des méthodes analytiques, deux vues sont considérées comme similaire si les primitives visuelles de celles-ci partagent la même topologie.

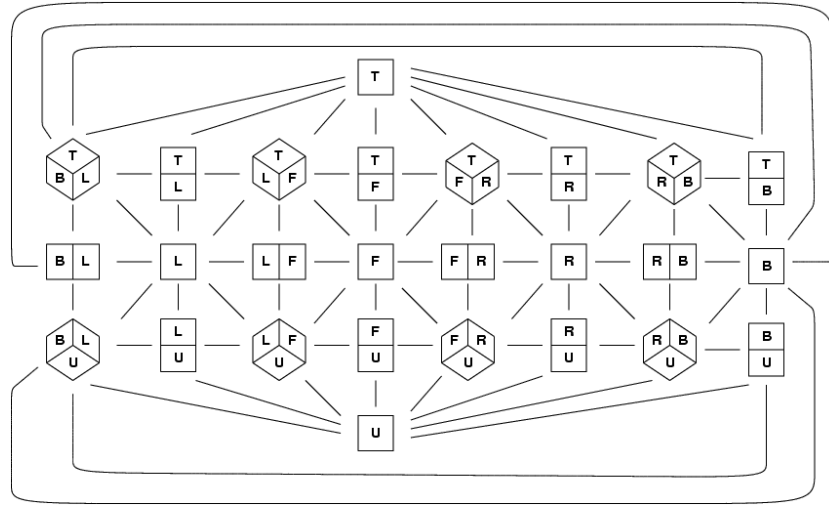


FIG. 1.1 – Graphe d'aspect pour un cube.

d'observer un objet depuis un point de vue accidentel est négligeable. Les aspects correspondant à ces points de vue sont donc généralement ignorés dans le modèle.

Cependant, les points de vue accidentels sont primordiaux lors de la création du modèle. En effet, les régions de la sphère de vues sont généralement établies à partir de leurs frontières. Il s'agit alors d'identifier les points de vue accidentels afin de déterminer les frontières de différentes régions. De telles solutions ont été proposées pour les objets polyédriques [29, 73] comme pour des objets "courbes" [66].

Méthodes empiriques

Les méthodes empiriques s'appuient sur l'hypothèse selon laquelle l'apparence d'un objet varie peu pour de petits déplacements. L'aspect d'une vue permet alors de caractériser l'ensemble des aspects pris par l'objet au voisinage de ce point de vue. Il suffit alors de réaliser un échantillonnage suffisamment dense de l'espace des poses pour obtenir un ensemble de points de vue couvrant les différents aspects de l'objet [13], cet échantillonnage pouvant être déterministe [7] ou aléatoire [19]. Afin de réduire la dimension de l'espace des poses à échantillonner, le modèle de caméra considéré correspond généralement un modèle de projection orthographique. Ainsi, une translation de la caméra le long de son axe optique ne modifie pas l'aspect de l'objet (à un facteur d'échelle près). L'objet peut alors être décrit par un ensemble de vues pris par une caméra se déplaçant sur la surface

d'une sphère centrée sur l'objet. La taille de cet ensemble peut néanmoins être relativement importante. Il peut alors être nécessaire de la réduire en fusionnant les vues pour lesquels l'objet partage un aspect très similaire[69].



FIG. 1.2 – Exemple de sphère de vues

Une fois l'ensemble de vues établi, il reste à décrire l'apparence de l'objet dans chacune de ces vues. On peut alors distinguer trois approches : les approches géométriques, les approches photométriques et les approches basées sur des descripteurs locaux.

1.2.2 Identification d'une vue à l'aide de primitives géométriques

Les approches géométriques caractérisent l'objet par sa forme. Puisque la forme d'un objet rigide est invariante quelque soit l'instance de l'objet considérée et les conditions d'illuminations de la scène, le seul paramètre dont dépend cette forme correspond à la pose de l'objet par rapport à la caméra. Celle-ci se manifeste à l'image au travers des contours de l'objet. Par conséquent, les primitives visuelles utilisées pour décrire l'aspect d'un objet sont donc des primitives géométriques telles que des courbes, des segments, des ellipses, etc.

Pour un point de vue donné, la forme d'un objet peut être décrite soit localement (l'objet est considéré comme une collection de primitives géométriques), soit globalement (l'objet est considéré comme une primitive géométrique unique et indivisible).

Approches locales

Les approches locales représentent l'aspect d'un objet comme une collection de primitives géométriques [13, 40, 69, 43], comme l'illustre la figure 1.3. Le principal intérêt de cette approche est d'être robuste aux occultations partielles de l'objet puisqu'une occultation n'affecte qu'une partie des primitives.

Afin d'assurer la répétabilité de leur extraction, les primitives considérées sont généralement simples : points d'intérêt, segments, contours, etc. La contrepartie de cette simplicité est un pouvoir discriminant relativement faible des primitives. La description de l'apparence d'un objet repose alors avant tout sur la configuration spatiale de ces primitives.

Un premier niveau consiste à identifier de petits groupes de primitives dont la configuration spatiale est aisément identifiable (critères de parallélisme, d'alignement, ...) [69, 50], comme l'illustre la figure 1.3. Ce processus, généralement désigné sous le terme de regroupement perceptif, permet d'élaborer des primitives plus discriminantes.

Un second niveau consiste à caractériser la configuration spatiale des primitives au niveau de l'objet entier. Une première solution consiste à caractériser cette configuration à l'aide d'invariants perspectifs. Cependant, seuls les objets plans et quelques classes d'objets simples admettent de tels invariants [67, 72, 16]. Une seconde solution consiste à décrire explicitement la configuration spatiale des primitives de l'objet. Les primitives de l'objet sont alors reliées entre elles par des contraintes géométriques 2D. Pour qu'une image soit identifiée comme similaire à l'une des vues du modèle, un ensemble de correspondances, entre les primitives de l'image et les primitives de la vue du modèle, vérifiant ces contraintes géométriques doit être établi [13, 68]. Différentes méthodes ont été développées pour rechercher un tel ensemble d'appariements. Celles-ci sont présentées dans la section 1.3.2 dans le cadre d'appariements 2D/3D mais peuvent être étendues au cadre d'appariements 2D/2D.

Approches globales

Les approches globales caractérisent l'aspect d'un objet par sa silhouette. Une seule primitive correspondant au contour externe de l'objet est ainsi utilisée pour caractériser l'aspect de l'objet. Pour détecter et recalibrer un objet dans une image, il est alors nécessaire d'extraire les silhouettes des objets présents dans la scène et de comparer celles-ci aux silhouettes correspondant aux différentes vues du modèle de l'objet.

La difficulté d'extraction d'une silhouette reste la principale limitation de cette

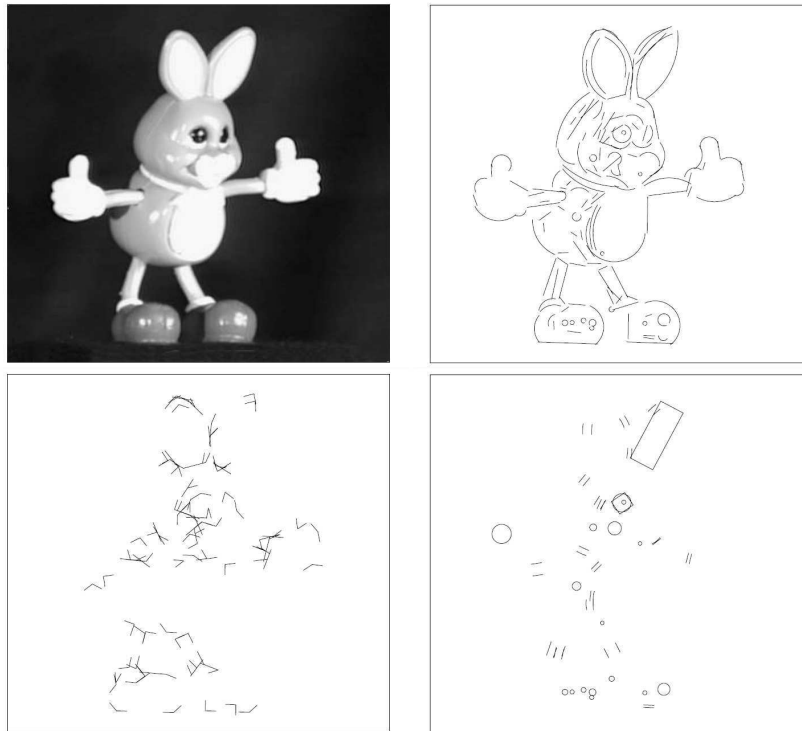


FIG. 1.3 – Exemple de représentation géométrique d’une vue par Pope et Lowe [69]. *L’image originale (en haut à gauche) est décrite sous la forme de courbes (en haut à droite), de L-jonctions (en bas à gauche) ainsi que de segments parallèles, d’ellipses et de régions (en bas à droite).*

approche. Il s’agit généralement d’un processus peu répétable excepté dans le cas d’objets isolés sur un fond uniforme. Néanmoins, comme le souligne Sethi et al. [78], lorsque des objets non-texturés sont courbes et présentent peu de discontinuités marquées à leur surface, la seule manifestation visuelle fiable correspond à la silhouette de l’objet.

Plusieurs approches ont été proposées pour caractériser une silhouette. Une première consiste à décrire la totalité de la silhouette. La silhouette est alors caractérisée à l’aide de transformées de Fourier [65], de transformées en ondelettes [42, 80], de coefficient de B-Spline [37], de moments [84, 28], d’analyse en composantes indépendantes [36], de modes de déformation [75], ...

D’autres approches caractérisent uniquement la configuration spatiale de points particuliers de la silhouette, comme les points d’inflexions et les points

de courbure extrême. La silhouette est alors caractérisée par la position de ces points, ceux-ci étant calculés pour différents lissages de la silhouette de manière à obtenir une représentation invariante à l'échelle [3, 60].

Enfin, certaines approches caractérisent la silhouette par sa structure [76, 52], celle-ci étant obtenue à l'aide d'opérateurs de morphologie mathématique. Puisqu'une occultation ne modifie pas la totalité de la structure de la silhouette, ces méthodes structurelles présentent généralement l'avantage d'être robustes aux occultations.

1.2.3 Identification d'une vue à l'aide de primitives photométriques

Les approches photométriques caractérisent l'aspect d'un objet, pour un point de vue donné, directement par l'image qu'il génère : *l'image modèle*. L'objet est alors identifié dans une image en comparant les valeurs des pixels de celle-ci avec celles des pixels de l'image modèle. En ne s'appuyant sur aucune primitive, ces approches évitent tout processus d'extraction de primitives et autres segmentations de l'image dont les résultats ne sont pas toujours fiables. En contrepartie, les approches photométriques ne disposent des informations de position, d'orientation et de facteur d'échelle que peuvent apporter les primitives des approches géométriques (e.g. une ellipse fournit une information de position, d'orientation et d'échelle). A la différence des approches géométriques, les approches photométriques ne peuvent donc pas exploiter ces informations pour réduire le nombre d'hypothèses de pose de l'objet à envisager. L'image modèle doit donc être comparée à l'image de la scène pour toutes les positions, orientations et échelles possibles. Dans certains cas, il est cependant possible de s'aider du contexte pour inférer certains de ces paramètres. Dans le cas d'une détection de piétons, par exemple, il est raisonnable de considérer que ceux-ci sont debout, ce qui revient à inférer leur orientation.

D'autre part, l'apparence photométrique d'un objet varie en fonction des instances de l'objet ainsi que des conditions d'illuminations. L'apparence d'un objet, même pour un seul point de vue, ne peut donc être décrit par une seule image. Une solution consiste à représenter un point de vue par autant d'images que nécessaire pour représenter la variabilité d'aspect de l'objet pour ce point de vue. Cette variabilité dépendant d'un grand nombre de paramètres (conditions d'illuminations, différences de texture et de couleur pour différentes instances de l'objet...), cet ensemble peut atteindre une taille très importante (plusieurs centaines d'images).

La taille du modèle devient alors rapidement trop importante pour que celui-ci soit stocké en mémoire. L'information contenue par ces ensembles d'images doit donc être représentée sous une forme plus concise.

Turk et Pentland [81], proposent de réduire la taille de cet ensemble de vues en représentant chacune des vues dans un même espace de faible dimension. En partant du constat qu'une grande part de l'information apportée par les images de cet ensemble est majoritairement redondante, ils proposent d'utiliser une Analyse en Composantes Principales (ACP) pour en extraire l'information principale sous la forme d'un petit nombre d'images caractéristiques : les *eigenimages* [81]. En effet, pour un ensemble de N images, l'ACP permet d'extraire N images caractéristiques (les vecteurs propres de l'ACP). L'ensemble des images représentable par une combinaison linéaire de ces images caractéristiques est désigné sous le terme d'*eigenspace*. Chacune des N images de l'ensemble de départ correspond à un point dans cet espace, c'est à dire que chacune d'elles peut être exprimée comme une combinaison linéaire de ces images caractéristiques. De plus, l'ACP ordonne les images caractéristiques de la manière suivante : les premières modélisent les principales variations d'aspect de l'objet alors que les dernières modélisent les variations les moins prépondérantes. Par conséquent, même en ne conservant que les M (avec $M \ll N$) premières images caractéristiques, il est possible d'approximer chacune des N images initiales à partir d'une combinaison linéaire de ces M images caractéristiques. L'ACP permet donc de passer d'images représentées dans un espace à très haute dimension à une représentation dans un espace à faible dimension.

Les *eigenimages* ont d'abord été utilisées pour la détection de visages : les *eigenfaces* [81]. Belhumeur et al. [5] ont proposé une variante des *eigenfaces* pour la reconnaissance de visages : les *fisherfaces*. Cette variante s'appuie non pas sur une ACP mais sur l'Analyse Discriminante de Fisher pour calculer la base d'images.

Murase et al. [61] ont repris le principe des *eigenfaces* pour l'adapter à la reconnaissance et le recalage d'objets 3D. Pour cela, un premier *eigenspace* global est utilisé pour différencier les différents objets. Cet *eigenspace* est calculé à partir de l'ensemble des images disponibles pour l'ensemble des objets. Dans cet *eigenspace*, un objet est représenté par une surface paramétrique. Cette surface est déterminée à partir de la projection des images de l'objet dans l'*eigenspace*. D'autre part, chaque objet dispose d'un second *eigenspace* personnel. Ce second *eigenspace* va permettre de représenter les variations d'aspect de l'objet en fonction de sa pose. A la différence du premier *eigenspace*, celui-ci est calculé à partir des seules images de l'objet. La projection des différentes images de l'objet dans ce second *eigenspace* permet de représenter l'objet par une surface paramétrée

par la pose de l'objet. Grâce à ces deux types d'*eigenspace*, on peut à la fois identifier l'objet ainsi que sa pose. L'identité est déterminée en projetant l'image de l'objet dans le premier *eigenspace* et en identifiant l'objet dont la surface est la plus proche de cette projection. Une fois l'objet identifié, la pose de l'objet est déterminée en projetant l'image de celui-ci dans l'*eigenspace* propre à cet objet. La pose de l'objet correspond alors aux paramètres associés au point de la surface représentant l'objet le plus proche de la projection de l'image. On notera que l'identification de l'objet et l'estimation de la pose nécessitent de fournir l'image de l'objet. Il est donc nécessaire d'avoir préalablement extrait l'objet du reste de l'image, tâche qui est généralement complexe. D'autre part, on notera que pour pouvoir projeter l'image de l'objet dans l'*eigenspace*, l'image de celui-ci doit avoir les mêmes dimensions que les *eigenimages*. L'image de l'objet doit donc être normalisée avant d'être projetée dans un *eigenspace*.

Afin de rendre les approches photométriques robustes aux occultations, Leonardis et al. [46] estiment la projection de l'image dans l'*eigenspace* de manière robuste. Cette projection est réalisée selon une approche d'*hypothèse/vérification*. Chaque hypothèse de projection est estimée à partir de quelques pixels extraits de l'image alors que l'approche classique utilise la totalité des pixels.

Huang et al. [35] proposent de remplacer l'approche globale des *eigenimages* par une approche locale dans laquelle l'objet est décomposé en parties représentées selon l'approche des *eigenimages*. Cette approche locale permet d'obtenir une robustesse aux occultations partielles. Cependant, elle réintroduit un processus de segmentation dont l'absence faisait la force des approches d'apparence photométrique.

Lee et al. [45] proposent une approche semi-globale. Ils remplacent l'ACP par une factorisation en matrice non négative (ou NMF) lors du calcul de la base d'images. Avec la NMF, les images caractéristiques sont de vraies images puisque la contrainte de non négativité interdit l'usage de pixel à valeur négative dans les images caractéristiques (ce qui n'est pas assuré avec l'ACP). En conséquence, les images caractéristiques présentent de très nombreux pixels nuls. Comme le montre la figure 1.4, les pixels non-nuls d'une image caractéristique sont généralement regroupés en un petit nombre de régions. Afin d'assurer cette propriété, Li et al. [48] proposent d'ajouter une contrainte de localité au processus de factorisation, cette dernière assurant que les pixels non-négatifs d'une image caractéristique soient bien regroupés. Il s'agit alors d'une factorisation locale en matrice non-négative ou LNMF. Puisque chacun des ensembles de pixels non-négatifs correspond à une petite partie de l'objet, la LNMF introduit implicitement une description de l'objet par partie dans une approche pourtant globale. Ce caractère local permet alors d'obtenir une meilleure robustesse aux occultations partielles.

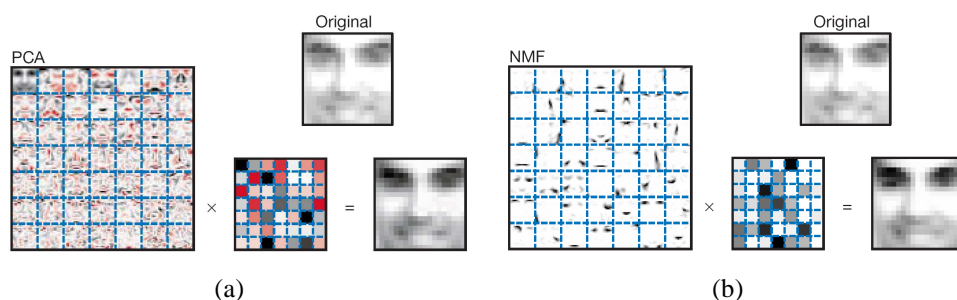


FIG. 1.4 – Comparaison entre l'utilisation de l'ACP et de la NMF pour la représentation de visages par Lee et al. [45]. Dans la figure (a), le visage "original" est approximé par une combinaison linéaire d'images caractéristiques issues d'une ACP et représentées dans la matrice de gauche. Les coefficients de cette combinaison linéaire sont représentés dans la matrice de droite et l'image-résultat correspond au visage placé à droite. La couleur rouge est utilisée pour représenter des valeurs négatives. De la même manière, la figure (b) présente l'approximation de l'image du même visage à partir d'images caractéristiques calculées par une NMF.

1.2.4 Identification d'une vue à l'aide de descripteurs locaux

Les approches basées sur des descripteurs locaux caractérisent l'apparence d'un objet à l'aide d'un ensemble de régions : les régions supports. Chacune de ces régions est identifiée à l'aide d'une signature calculée à partir du contenu de la région, ce contenu pouvant être géométrique comme photométrique. La description d'une vue correspond alors à la configuration spatiale de cet ensemble de régions signées. Les primitives employées dans ces approches sont donc des régions identifiées par leur contenu appelées *descripteurs locaux*.

Il s'agit de méthodes locales, donc robustes aux occultations. En ce qui concerne la stabilité et le pouvoir discriminant de ces approches, celles-ci dépendent principalement de la répétabilité de l'extraction des régions supports ainsi que du pouvoir discriminant et de l'invariance des signatures qui les caractérisent.

La région support

La qualité d'une région support repose principalement sur la répétabilité du processus d'extraction qui l'a définie. L'extraction d'une région doit être stable vis à vis :

- des conditions d'illuminations et du bruit dans l'image : les variations d'illu-

mination ainsi que la présence de bruit ne doivent pas ou peu modifier le résultat de l'extraction ;

- du point de vue : les régions extraites pour des points de vue différents doivent correspondre à la projection de la même surface 3D de l'objet. Les régions extraites doivent donc être covariantes avec les transformations appliquées au contenu de l'image comme l'illustre la figure 1.5. Ceci qui signifie qu'une région extraite d'une image ayant subi une transformation affine T correspond à la région support, extraite de l'image d'origine, à laquelle on applique la transformation T .

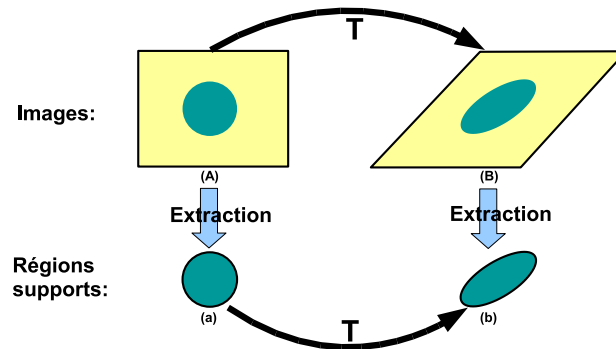


FIG. 1.5 – Principe de la covariance des régions extraites. Soient (A) et (B) deux images reliées par une transformation affine T . Soient (a) et (b) les régions support extraites de ces deux images. Si les régions extraites sont covariantes avec les transformations affines, alors (a) et (b) sont reliées par la transformation affine T .

Afin d'extraire des régions supports covariantes avec les similitudes, Lindeberg [49] propose d'utiliser une représentation multi-échelle de l'image : le *scale-space*. Pour détecter des "patatoïdes" (ou *blob*), le *scale-space* correspond à la réponse, pour chaque pixel de l'image, d'un filtre de *Laplacian-of-Gaussian* (ou LoG) et cela pour différentes tailles de gaussiennes. Le *scale-space* correspond alors à la réponse de la fonction LoG dans un espace tridimensionnel (x,y,taille). Afin d'obtenir des régions supports dont l'extraction est répétable, Lindeberg les définit à partir des maxima locaux de cette fonction dans le *scale-space* : les valeurs de x et y indiquent la position du centre de la région à l'image alors que la taille de la gaussienne définit le rayon de la région (la région considérée étant un disque).

Lowe [51] propose une implémentation rapide de l'approche de Lindeberg en approximant la fonction LoG par une différence de gaussienne (ou DoG) et en

calculant le *scale-space* à l'aide d'une pyramide d'images.

Mikolajczyk [56] propose de remplacer la fonction LoG par la réponse du détecteur de points d'intérêt de Harris [33]. De la même manière, les extremas de cette fonction permettent de définir une région support circulaire. Afin d'assurer la stabilité de la détection, seules les régions supports coïncidant avec un maximum de la fonction LoG dans le *scale-space* sont retenues. Un processus itératif, reposant sur l'étude des valeurs propres de la matrice de second ordre de la région support, permet ensuite de transformer cette région circulaire en une région elliptique. La région support ainsi définie présente l'avantage d'être covariante avec les transformations affines appliquées à l'image, ce qui signifie qu'une région extraite d'une image ayant subi une transformation affine A correspond à la région support, extraite de l'image d'origine, à laquelle on applique la transformation A .

Afin de ne pas construire le *scale-space* à partir de filtres répondant à un type particulier de primitives, Kadir [41] propose d'utiliser la mesure de l'entropie des niveaux de gris de l'image pour établir le *scale-space*. Les régions supports correspondent alors à des régions où l'entropie est maximale, ce qui correspond aux régions comportant le plus d'informations.

Mikolajczyk et al. [59] proposent d'utiliser la réponse du détecteur de contour de Canny [17] pour obtenir un *scale-space* représentant l'image de contour d'une scène. Chaque point de contour du *scale-space* est alors associé à la région circulaire centrée sur ce point pour laquelle la fonction LoG atteint un extremum local.

Parmi les approches se distinguant des précédentes, Matas et al. [53] proposent d'utiliser le processus "d'inondation" des méthodes de segmentation de l'image par ligne de partage des eaux (ou LPE) pour extraire les régions supports. D'une manière générale, une segmentation par LPE consiste à réaliser une segmentation binaire de l'image avec un seuil sur le niveau de gris (ou niveau d'inondation) de plus en plus élevé. Les régions de l'image dont l'intensité est inférieure à ce seuil forment alors des "bassins d'inondation". Au cours du processus, l'augmentation du niveau d'inondation entraîne un grossissement de ces bassins. Chaque bassin présente une vitesse d'expansion propre qui varie en fonction du niveau d'inondation. Matas et al. proposent d'identifier les bassins qui présentent une vitesse d'expansion particulièrement faible pour de nombreux niveaux d'inondation successifs. Pour un tel bassin, la région de l'image couverte lorsque sa vitesse d'expansion est faible définit alors une région support. Ces régions supports sont désignées sous le terme de *Maximally Stable Extremal Regions* ou MSER. Celles-ci présentent l'avantage d'être stables, robustes aux variations d'illumination et covariantes avec les transformations homographiques.

Sellinger [77] propose de définir une région support à partir de contours longs. La région support correspond alors à un carré centré sur le contour, orienté selon l'axe principal du contour, et dont la taille est définie à partir de la taille du contour le long de cet axe.

Mikolajczyk et al. [58] proposent une évaluation des performances de nombreux extracteurs de régions covariantes avec les transformations affines.

La signature

La signature d'un descripteur local doit caractériser l'information contenue dans la région. Elle doit être invariante aux variations d'illumination et aux transformations induites par un changement de point de vue. D'autre part, elle doit être discriminante afin de pouvoir différencier les différentes régions entre elles et ainsi permettre une mise en correspondance entre régions homologues issues de points de vue différents.

L'invariance de la signature vis à vis des changements de points de vue est assurée grâce à la propriété de covariance des régions supports avec les transformations induites par ces changements de point de vue. Pour deux régions supports correspondant à une même surface de l'objet observée selon deux points de vue différents, la covariance assure l'existence d'une transformation reliant les deux régions. Toutes les régions supports caractérisant une même surface de l'objet sont donc toutes équivalentes à une transformation près. Par conséquent, une représentation unique peut être utilisée pour représenter celles-ci : la région support normalisée. Dans le cas d'une région rectangulaire, cette normalisation consiste généralement à transformer la région en un carré unitaire. Dans le cas d'une région support circulaire ou elliptique, cette normalisation consiste à transformer la région support en un disque unitaire. Dans les deux cas, l'orientation de la région support normalisée doit être déterminée afin d'assurer l'unicité de la région support normalisée. Cette orientation est généralement déterminée à l'aide du contenu de la région support. Lowe [51] utilise la direction principale des gradients de la région alors que Obdrzalek et al. [64] recherchent un point extrémal pour fixer cette direction. Une autre solution consiste à caractériser le contenu de la région support normalisée à l'aide d'une signature invariante aux rotations afin de ne pas avoir à fixer cette orientation.

Le calcul de la signature consiste alors à caractériser le contenu de la région support normalisée. La signature peut alors caractériser l'apparence photométrique ou la forme de l'objet dans la région support normalisée.

Lazebnik [44] propose de caractériser la distribution normalisée des niveaux de gris en fonction de la distance au centre de la région. La région support normalisée est alors découpée en anneaux et un histogramme des niveaux de gris est établi pour chaque anneau. L'ensemble des niveaux de gris correspond alors à la signature de la région. Cette signature est désignée sous le terme de *spin image*.

Pour réduire la sensibilité de la signature aux variations d'illumination, Lowe [51] propose de caractériser le contenu de la région support par la distribution de l'orientation du vecteur gradient en fonction de la position du pixel considéré dans la région. Afin d'obtenir une signature de taille raisonnable et d'accroître la robustesse de la signature aux erreurs commises lors de l'extraction de la région, cette distribution est quantifiée. Ainsi, la région est découpée en baquets (généralement une grille 4×4) et la distribution des orientations des gradients est calculée sur l'ensemble des pixels du baquet. Cette distribution des orientations est, elle aussi, quantifiée puisque l'espace angulaire est découpé en un nombre fini de secteurs angulaires (généralement 8 secteurs). Cette signature est généralement désignée sous le terme *SIFT*.

Dans le cas d'objets peu texturés, Belongie [6] propose de caractériser la distribution spatiale de contours. La région normalisée est alors découpée en baquets à l'aide d'une grille radiale et le nombre de points de contour contenus dans chacun des baquets est stocké dans une matrice. Cette matrice correspond à la signature de *Shape context*.

De nombreuses autres approches ont été proposées pour caractériser le contenu d'une région support. Une description ainsi qu'une évaluation d'une grande partie d'entre elles sont proposées par Mikolajczyk et Schmid [57].

1.3 Approches 2D/3D

A la différence des approches 2D/2D, la connaissance a priori exploitée par les approches 2D/3D n'est pas celle des différents aspects de l'objet mais celle de la forme 3D de celui-ci. L'objet est alors représenté par son *modèle géométrique 3D*, c'est à dire un ensemble de primitives mathématiques 3D dont la configuration spatiale est exprimée dans un repère unique (le repère de l'objet).

A la différence d'un modèle d'apparence, un modèle géométrique ne décrit pas ce que l'on observe à l'image mais la structure 3D qui génère ces manifestations visuelles. Le processus de recalage consiste alors à déterminer la pose permettant d'aligner les primitives 3D du modèle avec les manifestations qu'elles génèrent à l'image (cf. figure 1.6). A l'aide d'une modélisation du processus de génération de l'image (i.e. choix du modèle de caméra et calibrage de celle-ci), il devient

possible de calculer cette pose à partir de quelques mises en correspondances entre les primitives du modèle et celles de l'image [20, 34, 23].

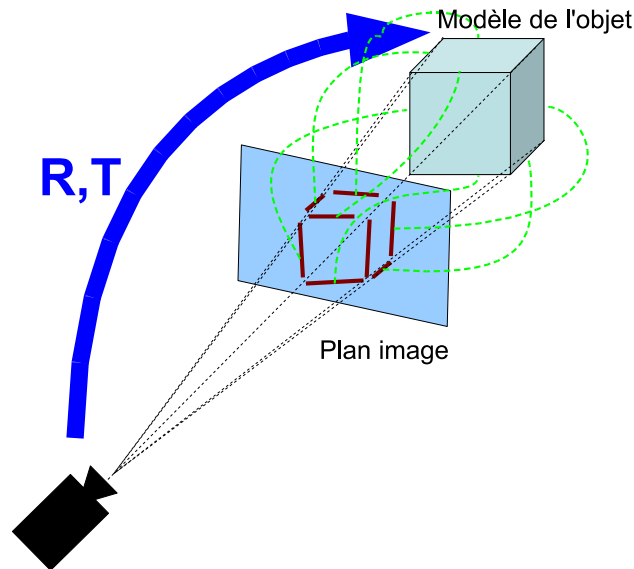


FIG. 1.6 – Principe des approches 2D/3D. Les appariements (pointillés verts), entre les primitives visuelles de l'image (en rouge) et les primitives du modèle, sont utilisés pour calculer la pose $[R|T]$ de l'objet.

Cependant, cette approche se heurte à deux difficultés. Tout d'abord, le modèle de l'objet ne peut être établi en ignorant le processus d'extraction des primitives visuelles. Le modèle doit être élaboré à partir de primitives 3D générant à l'image des primitives visuelles dont l'extraction est stable et répétable (ce point est détaillé dans le paragraphe 1.3.1).

D'autre part le processus de "mise en correspondance/calcul de la pose" doit assurer la contrainte de consistance du point de vue (ou *viewpoint consistency*), ce qui signifie que toute primitive de l'image doit correspondre à la projection, selon la pose calculée, de la primitive du modèle à laquelle elle est appariée. Ce point est abordé dans le paragraphe 1.3.2.

1.3.1 Choix des primitives du modèle

L'inconvénient majeur des approches exploitant un modèle géométrique est qu'il n'est généralement pas possible de comparer directement les primitives du modèle à celles issues de l'image, ceci en raison de leur nature différente. En effet,

les primitives de l'image correspondent à la projection dans l'image des primitives du modèle. Pour pouvoir les comparer, les approches 2D/3D doivent inférer les manifestations visuelles correspondant aux projections des primitives du modèle. Or, il est généralement difficile d'inférer la manifestation visuelle d'une primitive 3D quelconque lorsque l'on ignore le point de vue depuis lequel celle-ci est observée. De plus, la manifestation visuelle doit correspondre à une primitive visuelle dont l'extraction est stable et répétable.

Par conséquent, les primitives utilisées pour établir le modèle sont généralement des primitives 3D simples dont la projection à l'image est prédéterminée et facile à extraire. C'est en particulier le cas des points 3D ou des segments 3D dont les primitives visuelles correspondantes sont respectivement des points d'intérêt [33] et des segments de contours [15].

1.3.2 Mise en correspondance et estimation de la pose

Si l'utilisation de primitives 3D simples permet d'extraire facilement de l'image les manifestations visuelles qui leurs correspondent, le faible pouvoir discriminant de ces primitives ne permet pas d'envisager un processus de mise en correspondance considérant les primitives indépendamment.

Ce pouvoir discriminant peut néanmoins être améliorée en formant des ensembles de primitives ayant une configuration spatiale identifiable. Afin de pouvoir facilement déterminer les groupes de primitives visuelles qui leurs correspondent, les primitives 3D sont regroupées de manière à ce que celles-ci génèrent des primitives visuelles présentant une configuration spatiale remarquable et quasi-invariante aux changements de points de vue (alignement, parallélisme, ...) [50]. La théorie de la Gestalt est généralement utilisée pour identifier ces ensembles de manifestations visuelles, cette technique étant désignée sous le terme de regroupements perceptifs.

Si ces derniers permettent de réduire le nombre d'appariements envisageables, ils ne permettent pas d'assurer une mise en correspondance unique consistante avec un point de vue. Pour obtenir une mise en correspondance correcte, le nombre d'appariements consistants avec un point de vue doit être maximal. Trois familles d'approches ont été proposées pour réaliser cette tâche :

- les approches travaillant dans l'espace des appariements : celles-ci identifient d'abord la mise en correspondance en explorant l'espace des appariements à l'aide de contraintes géométriques, la pose de l'objet étant ensuite calculée à partir de ces appariements ;
- les approches travaillant dans l'espace des poses : celles-ci identifient tout d'abord la pose de l'objet puis déterminent la mise en correspondance à

- partir de cette pose ;
- les approches d'alignements : celles-ci exploitent une méthode dite *d'hypothèse/vérification* afin d'explorer "simultanément" l'espace des poses et l'espace des appariements.

Méthodes travaillant dans l'espace des appariements - Arbres d'interprétations

L'**espace des appariements** correspond à l'ensemble des appariements envisageables entre les primitives issues du modèle et les primitives issues de l'image. Afin de pouvoir représenter les primitives visuelles ne correspondant pas à une manifestation visuelle de l'objet, la primitive *nil* est ajoutée aux primitives du modèle. Ainsi, chaque primitive de l'image est associée à un ensemble d'hypothèses d'appariements. Lorsqu'une primitive visuelle ne correspond à aucune primitive du modèle, celle-ci est représentée dans l'espace des appariements par un unique appariement avec la primitive *nil*. Par contre, une primitive visuelle pouvant correspondre à N primitives du modèle est associée à $N + 1$ hypothèses d'appariement, soit un appariement avec chacune des N primitives du modèle plus l'appariement avec *nil*. Grimson a proposé d'organiser l'ensemble des hypothèses d'appariements sous la forme d'un arbre : l'*arbre d'interprétation* [31]. Chaque niveau de l'arbre correspond au choix d'une hypothèse d'appariement pour une primitive visuelle. Ainsi, chaque chemin de cet arbre définit une mise en correspondance. La mise en correspondance optimale correspond alors au chemin passant par le moins de noeud contenant un appariement avec *nil* tout en définissant un ensemble d'appariements consistant avec un point de vue.

Grimson a proposé d'utiliser un parcours en profondeur de l'arbre pour identifier cette mise en correspondance optimale. Afin d'accélérer ce parcours, des contraintes géométriques locales peuvent être employées pour déterminer les branches qui ne peuvent pas aboutir à une solution consistante et ainsi élaguer l'arbre [14, 31]. Les contraintes locales n'assurant pas nécessairement que la solution soit consistante avec un point de vue, une étape de vérification du résultat reste nécessaire. Cette vérification passe généralement par l'estimation de la pose de l'objet et la vérification de la cohérence de la mise en correspondance vis à vis de cette pose.

Méthodes travaillant dans l'espace des poses

Ces méthodes ne déterminent pas explicitement une mise en correspondance. Le principe est de déterminer, pour chaque point de l'espace des poses, le nombre

d'appariements consistants avec ce point de vue. La pose, pour laquelle le nombre d'appariements consistants est maximal, est alors considérée comme la solution du recalage de l'objet.

Une des méthodes travaillant dans l'espace de pose est la Transformée de Hough Généralisée [4]. Celle-ci associe un score à chaque point de l'espace des poses. Initialement, ce score est nul pour tout point de cet espace. Ensuite, pour chaque appariement, toutes les poses consistantes avec cet appariement sont déterminées et le score de celles-ci est incrémenté d'une unité. La pose ayant le score final le plus élevé correspond alors à la pose de l'objet. Afin de pouvoir représenter l'espace des poses en mémoire, celui-ci est discrétisé plus ou moins grossièrement. Par conséquent, le vote d'un appariement n'est plus attribué à une pose précise mais à une région de l'espace des poses. De ce fait, plus la taille de ces régions est importante, moins il est nécessaire d'estimer les poses cohérentes avec précision puisqu'il suffit que celles-ci appartiennent à la bonne région de l'espace des poses. La discrétisation de l'espace des poses améliore donc la robustesse de l'algorithme aux erreurs sur l'estimation des poses cohérentes avec un appariement. Cependant, cette discrétisation ne permet plus d'assurer la consistance des appariements ayant voté pour une même région de l'espace puisque les votes sont généralement issus de poses différentes. Si ce problème est négligeable dans le cas d'une discrétisation fine de l'espace des poses (les poses d'une même région sont suffisamment similaires pour que les appariements ayant votés soient consistants), ce n'est plus le cas pour une discrétisation grossière. D'autre part, la précision de la pose obtenue à l'issue du processus dépend de la finesse de ce découpage.

Afin d'éviter ce découpage arbitraire de l'espace des poses, Cass [18] et Breuel [13] ont proposé un découpage adaptatif. Afin de prendre en compte l'incertitude de localisation de la primitive image, l'erreur sur la pose consistante associée est bornée. Ainsi, chaque pose consistante est remplacée par un petit volume de l'espace des poses. Par conséquent, l'ensemble des poses consistantes avec un appariement définit un volume de l'espace des poses. L'ensemble des appariements génère alors un ensemble de volumes. La pose de l'objet correspond alors à la zone de l'espace correspondant à l'intersection du plus grand nombre de ces volumes. La recherche de cette zone peut être accélérée à l'aide de techniques de *branch and bound*. Cette approche, initialement présentée pour des objets plans, a été étendue par Dementhon [20] pour des objets non plans.

Méthodes d'alignement

Les méthodes d'alignement [39, 38, 27] travaillent à la fois dans l'espace des appariements et dans l'espace des poses. Tout d'abord, un ensemble d'appariements permettant de calculer une pose est extrait de l'espace des appariements (la taille de cet ensemble est généralement minimale). Cet ensemble d'appariements permet donc de calculer une pose, soit un point dans l'espace des poses. Tout comme pour les méthodes travaillant dans l'espace des poses, le nombre d'appariements cohérents avec cette pose est déterminé. Celui-ci est obtenu en projetant les primitives du modèle dans l'image selon cette pose et en déterminant les appariements cohérents à l'aide de leur erreur de reprojection.

Les appariements ayant servi au calcul de la pose pouvant être erronés, ce processus est répété un grand nombre de fois. A chaque itération, un point de l'espace des poses est étudié, ce qui revient à effectuer un échantillonnage de l'espace des poses. A l'issue de cet échantillonnage, la pose cohérente avec le plus grand nombre d'appariements est retenue comme solution. Le nombre d'itérations à effectuer est généralement déterminé à partir de la proportion de faux appariements ainsi que de la probabilité de réussite souhaitée de l'algorithme.

1.4 Approches 3D/3D

Tout comme les approches 2D/3D, les approches 3D/3D exploitent la connaissance a priori de la forme 3D de l'objet. Par conséquent, les approches 3D/3D ne peuvent pas, elles aussi, comparer directement les primitives 3D de leur modèle aux primitives 2D de l'image. Alors que les approches 2D/3D résolvent ce problème en inférant l'apparence à l'image des primitives 3D, les approches 3D/3D utilisent la solution duale puisqu'elles infèrent la géométrie 3D des primitives visuelles issues de l'image. Ainsi, les approches 3D/3D élaborent un modèle géométrique 3D de la scène pour ensuite pouvoir réaliser une comparaison entre celui-ci et le modèle de l'objet (cf. figure 1.7). Le recalage consiste alors à déterminer la transformation rigide permettant d'aligner au mieux les deux modèles 3D. Il s'agit alors d'un processus purement 3D ne faisant ni intervenir la pose de la caméra ni aucun processus de projection à l'image. En particulier, il n'est pas nécessaire de tester la consistance d'un appariement pour un point de vue, le respect des contraintes géométriques 3D du modèle de l'objet étant suffisant pour assurer la cohérence de l'appariement.

Cependant, inférer la géométrie d'un objet, ou de ses primitives, à partir d'une seule vue est un problème complexe ne pouvant être résolu qu'à l'aide de connais-

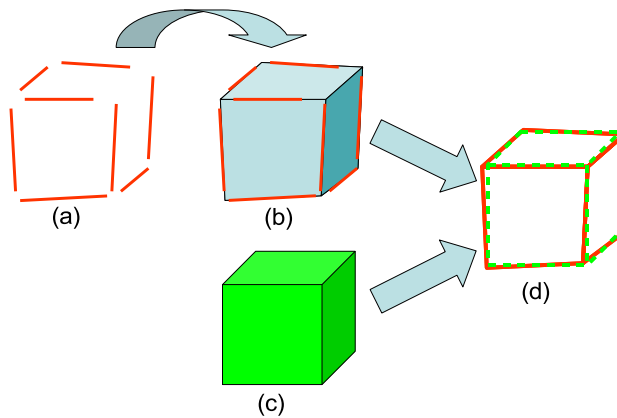


FIG. 1.7 – Approche 3D3D. (a) : les segments extraits d'une image. (b) : reconstruction du modèle géométrique de la scène observée. (c) : modèle de l'objet. (d) : comparaison du modèle de la scène et du modèle de l'objet.

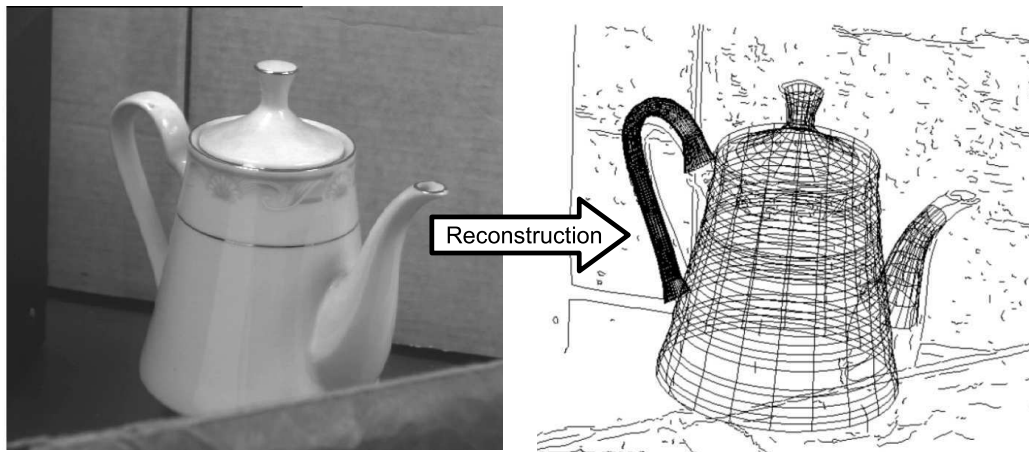


FIG. 1.8 – Exemple d'une reconstruction 3D à l'aide de cylindres généralisés obtenue par Medioni et al. [55]. A gauche : l'image originale. A droite : les contours extraits de l'image superposés avec la représentation en fils de fer du modèle 3D reconstruit de la scène.

sances a priori. Pour y parvenir, ces approches restreignent la classe d'objets considérée. Ainsi, Roberts [70] propose un algorithme permettant de reconstruire, reconnaître, et finalement localiser des objets polyédriques. Dickinson et al. [24] proposent de considérer uniquement des objets pouvant être décomposés en un ensemble de primitives simples : les *geons*. Nevatia et al. proposent une solution

pour certains types de cylindres généralisés [62] alors que Medioni et al. considèrent la classe des objets composés de cylindres généralisés [55] (cf. figure 1.8).

Si la possibilité d'exploiter un modèle 3D de la scène présente d'intéressantes perspectives, en particulier la possibilité d'exploiter directement les contraintes de rigidité 3D reliant les primitives du modèle pour déterminer un ensemble d'appariements cohérent, ces approches sont principalement limitées par la complexité du processus d'élaboration du modèle de la scène observée ou par la trop grande restriction sur la nature des objets considérés.

1.5 Approches mixtes

On regroupe sous le terme *approches mixtes* l'ensemble des approches combinant au moins deux des trois approches précédemment exposées. Celles-ci combinent généralement une procédure de mise en correspondance issue des approches 2D/2D, en raison de leur rapidité, avec une méthode de calcul de pose 2D/3D permettant d'atteindre un recalage précis. Elles reposent par conséquent sur un modèle combinant à la fois une description de l'apparence et une description de la géométrie 3D de l'objet. Celui-ci est généralement obtenu en déterminant la position 3D de primitives issues d'un modèle d'apparence et en organisant celles-ci sous la forme d'un modèle géométrique. Plusieurs solutions ont été proposées pour élaborer un tel modèle.

Alleazard et al. [1] proposent de générer le modèle d'apparence à partir du modèle 3D texturé de l'objet. Le modèle d'apparence est établi, à partir de vues de synthèses, selon une approche basée sur des descripteurs locaux de texture. Ces descripteurs étant basés sur des points d'intérêt de Harris, la position 3D associée à chaque descripteur correspond à celle du point 3D du modèle se projetant sur le point d'intérêt. La maîtrise complète du processus de génération des images permet de déterminer cette position 3D.

La solution duale consiste à reconstruire un modèle 3D à partir d'une sphère de vues réelles. Ainsi, Rothganger et al. [71] reconstruisent la position 3D de chaque descripteur local à partir des observations dans différentes vues de la région support.

Dans les deux cas, le modèle final associe une information géométrique 3D à chaque descripteur local. Le modèle 3D n'est donc pas constitué de primitives mathématiques abstraites et difficiles à apparier avec les primitives visuelles, mais de primitives directement observables dans l'image. Ainsi, le processus de mise en correspondance des primitives du modèle avec celles de l'image est similaire à ceux employés dans les approches 2D/2D et la pose est calculée à partir des

appariements 2D/3D. Ces approches mixtes permettent donc d'allier la rapidité et le pouvoir discriminant des mises en correspondance 2D/2D avec la précision et la robustesse des calculs de pose 2D/3D.

1.6 Position des travaux présentés

Chacune des approches 2D/2D, 2D/3D ou 3D/3D présente ses avantages et ses inconvénients. Les approches 2D/3D permettent d'obtenir des recalages fins grâce au modèle géométrique 3D de l'objet. Cependant, la différence de nature entre les primitives de ce modèle et les primitives visuelles extraites de l'image rend leur mise en correspondance difficile. A l'opposé, les approches 2D/2D utilisent un modèle d'apparence qui leur permet d'établir une mise en correspondance rapide. Cependant, l'absence d'un modèle tridimensionnel limite la précision des poses estimées. En ce qui concerne les approches 3D/3D, leur capacité à élaborer un modèle géométrique 3D de la scène leur permettent d'exploiter directement les contraintes géométriques 3D du modèle de l'objet afin d'identifier les groupes de primitives cohérentes avec ce même modèle. En travaillant directement dans l'espace 3D (aussi appelé espace objet), ces méthodes évitent de faire intervenir le processus de projection, ce dernier présentant généralement l'inconvénient de nécessiter des ajustements en fonction de la distance d'observation. Néanmoins, ces méthodes 3D/3D nécessitent d'inférer la géométrie 3D sous-jacente à une primitive visuelle. Même en limitant la classe d'objets étudiés (objets constitués de cylindres généralisés, objets polyédriques, ...), cette étape reste généralement délicate.

Pour pallier à ces difficultés, les approches mixtes tentent de combiner ces différentes approches, le plus souvent l'approche 2D/2D avec l'approche 2D/3D. Il s'agit alors d'exploiter la capacité à identifier les primitives de l'image des approches 2D/2D ainsi que la capacité à calculer une pose précise des approches 2D/3D.

Ces approches mixtes présentent donc de nombreux avantages. Cependant, les solutions proposées jusqu'à présent sont uniquement adaptées aux objets texturés. Celles-ci ne permettent donc pas de traiter des objets complexes peu texturés comme le sont généralement les objets mécaniques métalliques. La solution proposée dans cette thèse correspond donc à une approche mixte adaptée aux pièces mécaniques complexes, généralement peu texturées et comportant des alésages.

D'autre part, un simple processus de mise en correspondance 2D/2D n'est généralement pas suffisant discriminant lorsque les pièces étudiées comportent de

nombreux motifs répétitifs, c'est à dire des régions de l'objet partageant la même apparence mais pas la même position 3D. Or, de tels motifs répétitifs sont présents en grand nombre dans les pièces mécaniques. Nous proposons donc de compléter le processus de mise en correspondance 2D/2D par un processus de filtrage exploitant la contrainte de rigidité du modèle de l'objet afin d'assurer la cohérence des appariements établis. Une solution de filtrage exploitant une approche 2D/3D ainsi qu'une solution exploitant une approche 3D/3D sont proposés. Comme il a déjà été indiqué, la principale limitation des approches 3D/3D correspond à l'étape de reconstruction de la scène, c'est à dire l'étape au cours de laquelle la géométrie 3D d'une primitive visuelle doit être inférée. Or, en proposant un modèle d'objet combinant apparence et géométrie, ainsi qu'un processus permettant de calculer la pose de chaque primitive visuelle appariée au modèle, la solution présentée dans ce manuscrit permet d'établir un modèle géométrique 3D de la scène sans avoir à réaliser une reconstruction de celle-ci. Ce modèle peut alors être utilisé pour identifier les appariements cohérents selon une approche 3D/3D.

Chapitre 2

Modèle combinant apparence et 3D

Objectif : Pour pouvoir identifier un objet dans une image et inférer sa pose à partir de celle-ci, un système de vision doit disposer d'une représentation interne de cet objet : le modèle de l'objet. Ce chapitre présente le modèle que nous avons défini pour caractériser des pièces mécaniques métalliques. Celui-ci combine un modèle d'apparence à un modèle géométrique 3D.

Dans ce chapitre, un modèle combinant apparence et géométrie 3D est proposé. Ce modèle est élaboré à partir du modèle CAO de l'objet ainsi que d'une sphère de vues préalablement recalées avec le modèle CAO, comme l'illustre la figure 2.1. Des primitives visuelles issues des différentes vues sont associées à des primitives 3D et organisées en un modèle géométrique 3D. Les primitives visuelles sont présentées dans la section 2.1 alors que l'élaboration du modèle géométrique à partir de celles-ci est présentée dans la section 2.2. Enfin, une extension du modèle permettant d'estimer la pose d'une primitive visuelle à partir de son seul appariement à une primitive de modèle est proposée (section 2.3). Cette extension du modèle sera particulièrement utile lors de l'étape de filtrage des appariements présentée dans le chapitre 4.

Le modèle, tel que présenté dans ce chapitre, vise à répondre au contexte particulier de cette thèse. Ainsi, un certain nombre de choix se justifie par la nature des pièces étudiées, c'est à dire des pièces mécaniques peu texturées comportant des alésages. Dans un contexte différent, ces choix pourraient être remis en cause afin d'adapter le modèle à la nature des pièces considérées.

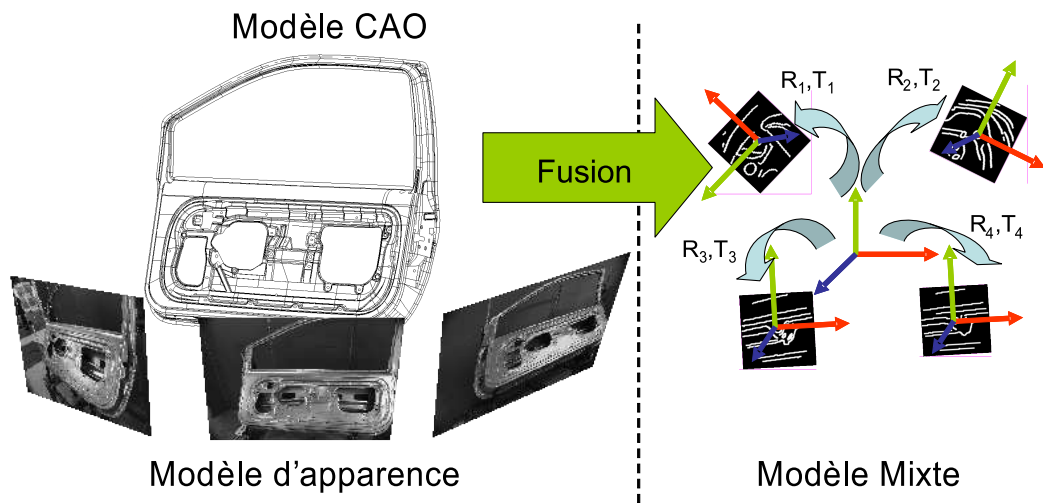


FIG. 2.1 – Principe du modèle combinant apparence et géométrie 3D. A gauche, le modèle géométrique est illustré par un maillage en "fil de fer" d'une portière alors que le modèle d'apparence est représenté par les 3 vues d'une sphère de vue. A droite, le modèle mixte associe chaque primitive visuelle (les imagerie de contours) à une primitive 3D (les repères). Ces primitives sont reliées par des contraintes géométriques 3D (les rotations R_i et les translations T_i).

2.1 Les primitives visuelles : des descripteurs locaux de contours

Plusieurs solutions existent pour modéliser l'apparence d'un objet : certaines modélisent son apparence photométrique alors que d'autres modélisent sa forme (cf. 1.2). Le contexte de la thèse, et plus particulièrement la nature des pièces à recalcr, a motivé le choix d'une description de la forme de l'objet reposant sur des descripteurs locaux de contours.

Les objets considérés dans cette étude correspondent à des objets mécaniques, généralement complexes et le plus souvent métalliques. Ils sont généralement peu texturés ou présentent une texture peu répétable d'une instance de l'objet à une autre. Pour illustrer ces propos, considérons le cas d'une pièce mécanique de fonderie telle qu'une culasse de moteur (cf. figure 2.2). Une telle pièce présente des surfaces brutes ainsi que des surfaces usinées. S'agissant des parties brutes, la texture présente à la surface correspond aux aspérités du moule. Deux pièces issues de deux moules différents ne présentent donc pas la même texture : la texture est

instable. S'agissant des parties usinées, les procédés tels que le polissage tend à supprimer toute texture de la surface de l'objet : ces surfaces ont généralement un aspect uniforme. Enfin, la nature métallique de l'objet implique la présence de fortes réflexions spéculaires accentuant l'instabilité de la texture de l'objet.

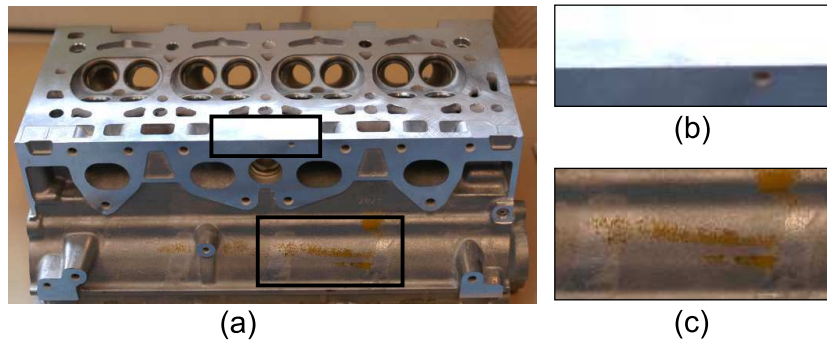


FIG. 2.2 – Instabilité de la texture des pièces mécaniques. *Une culasse de moteur (a) possède des zones polies ne présentant pas de texture (b) ainsi que des zones brutes dont la texture est liée aux aspérités du moule (c).*

A la différence de la texture, la forme d'un objet mécanique peut être considérée comme stable¹. Les manifestations visuelles de la forme d'un objet correspondent aux contours de celui-ci dans l'image, primitive relativement stable vis à vis des conditions d'illumination.

Le contexte de ces travaux n'excluant pas la présence d'objets parasites et ni d'occultations partielles, des descripteurs locaux de contours sont employés pour caractériser la forme de l'objet. Ces derniers présentent l'avantage d'être quasi-invariants² aux changements de point de vue tout en étant des primitives discriminantes.

Dans les paragraphes suivants, un descripteur local de contour, adapté au contexte de ces travaux, est proposé. Un tel descripteur étant défini par le choix d'une région support et d'une signature caractérisant cette région, ces deux points sont respectivement abordés dans les paragraphes 2.1.1 et 2.1.2.

¹Il est raisonnable de supposer que le procédé de fabrication assure le respect de la même géométrie pour l'ensemble des instances de l'objet produites à une précision compatible avec les traitements souhaités.

²Généralement, les descripteurs locaux ne sont pas invariants aux changements de points de vue mais à une approximations de ceux-ci (généralement sous la forme d'une transformation affine).

2.1.1 Région support : le voisinage d'un contour fermé

Comme présenté dans l'état de l'art, un descripteur local correspond à une petite région de l'image à laquelle une identité, définie à partir de son apparence, est assignée. La description du contenu d'une image, et a fortiori la description de l'image d'un objet, consistent alors à extraire de l'image les régions présentant l'information pertinente (les régions supports) et à caractériser chacune de ces régions.

Pour que cette description soit pertinente, il est souhaitable que les régions supports soient :

- nombreuses : plus grand est le nombre de régions décrivant l'objet, plus la description est robuste aux occultations partielles de l'objet ;
- répétables : en dépit de variations des conditions extérieures (conditions d'illumination, environnement), les régions extraites doivent être au moins pour partie les mêmes afin d'assurer une représentation de l'objet invariante à ces paramètres externes,
- covariantes avec les transformations induites par les changements de point de vue : les régions extraites selon deux points de vue différents d'un même objet doivent correspondre à la projection de la même surface de l'objet, ceci afin d'assurer le même contenu sémantique à l'intérieur de ces régions.

Parmi les extracteurs de régions répondant à ces contraintes, les plus utilisés sont sans doute les extracteurs dits "Laplace-affine" et "DoG-Affine". En effet, ces extracteurs de régions sont employés dans de nombreux systèmes de reconnaissance et de recalage d'objets [51, 44]. Cependant, dans le cas d'objets métalliques, les régions supports extraites par ces derniers sont peu répétables en raison des nombreuses réflexions spéculaires qui perturbent le processus d'extraction. La figure 2.3, présentant les résultats d'extraction de région par un DoG-Affine³ sur une de nos pièces, illustre cette instabilité. Dans ces images, les ellipses jaunes correspondent aux régions supports. Bien que les deux images correspondent à deux points de vue proches d'un même objet, celles-ci partagent très peu de régions supports communes. On peut noter que de nombreuses régions supports correspondent à des spécularités. La position et la forme de ces spécularités dépendant à la fois de la pose de la caméra et des conditions d'illumination, l'extraction de ces régions est très peu répétable.

Pour définir une région support adaptée aux pièces mécaniques, la solution proposée s'appuie sur la présence en grand nombre d'alésages dans ce type d'ob-

³Les régions supports ont été extraites à l'aide de l'implémentation du DoG-Affine de G. Dorko. Les exécutable permettant de réaliser cette extraction sont disponibles sur le site <http://lear.inrialpes.fr/people/dorko/downloads.html>

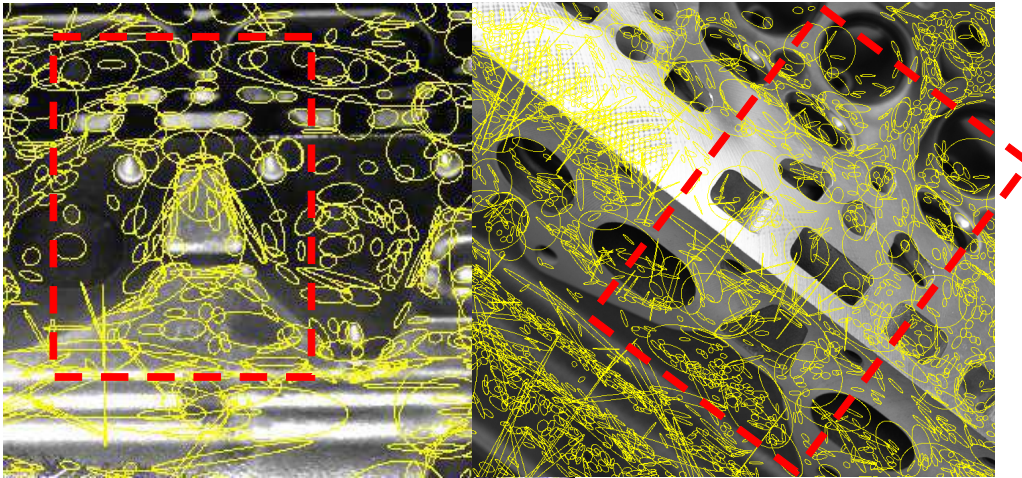


FIG. 2.3 – Région extraites par un DoG-Affine sur une pièce mécanique. *Les ellipses jaunes correspondent aux régions supports extraites par un Dog-Affine. Les images correspondent à deux points de vue différents d'un même objet mécanique (une culasse). Les cadres rouges correspondent à la même région de l'objet vu selon deux points de vue différents.*

jet. Les discontinuités géométriques associées à ceux-ci sont généralement suffisamment importantes pour générer des contours fermés dont l'extraction présente une répétabilité acceptable. Les régions supports extraites correspondent donc au voisinage de contours fermés.

Extraction des contours fermés

L'approche utilisée pour extraire des contours fermés consiste à extraire tous les contours de l'image puis à sélectionner ceux qui correspondent à des contours fermés. Pour augmenter la productivité de cette étape, quelques mécanismes sont ajoutés afin de favoriser l'extraction de contours fermés.

L'extraction des contours est réalisée par l'algorithme de Canny-Deriche [22]. Celui-ci se déroule en 3 étapes : tout d'abord le calcul de l'image des gradients, puis une étape de seuillage par hystérésis ne conservant que les pixels de gradient élevé, et enfin l'étape de sélection qui ne conserve que les pixels présentant un maxima du module du gradient dans la direction du gradient. C'est au cours des deux dernières étapes que les contours longs peuvent être fragmentés.

Afin d'éviter la suppression de pixels de contours dans une zone de faible contraste au cours de l'étape seuillage par hystérésis, une égalisation locale des ni-

veaux de gris est appliquée lors du calcul du gradient. L'algorithme utilisé est semblable à l'algorithme de CLAHE (Contrast Limited Adaptive Histogram Equalization) [85]. Celui-ci réalise une égalisation locale des niveaux de l'image tout en limitant le rehaussement du contraste, limitation qui permet de ne pas accroître de manière trop significative le bruit de l'image.

Au cours de l'étape de suppression des contours non maximaux, la fragmentation des contours survient généralement au niveau des coins (zones de forte courbure du contour) ainsi qu'au niveau des zones de l'image où l'orientation des gradients est bruitée. La solution proposée consiste à régulariser préalablement l'orientation de ces gradients. Cette régularisation est obtenue en remplaçant l'orientation de chaque pixel de contour par l'orientation principale du gradient sur un voisinage de ce pixel. Cette orientation est calculée à partir de l'ensemble des vecteurs gradients du voisinage à l'aide d'une Analyse en Composante Principale. L'orientation des gradients étant uniquement utilisée lors de l'étape de sélection, il suffit d'appliquer l'étape de régularisation juste avant cette dernière. De plus, puisque cette étape ne traite que les pixels sélectionnés par l'étape d'hystérésis, on peut restreindre l'application de la régularisation à ces seuls pixels, ce qui limite l'impact sur les performances de l'extracteur.

Afin d'accroître le nombre de contours fermés, des heuristiques de regroupements perceptifs sont employées. Une technique employée consiste à clôturer les contours quasi-fermés, c'est à dire ceux dont les extrémités sont proches les unes des autres vis à vis de la taille du contour⁴.

Enfin, cette extraction de contour est réalisée en multi-résolution, permettant ainsi l'extraction de structures à différentes échelles.

Définition de la région support

La région support associée à un contour fermé correspond à l'ellipse d'inertie [54] de ce contour à laquelle un facteur d'échelle est appliqué.

L'ellipse d'inertie \mathcal{E} d'un contour fermé Γ est obtenue à partir de la matrice de covariance $Cov(\Omega)$ de la région Ω délimitée par Γ :

$$Cov(\Omega) = \frac{1}{S(\Omega)} \sum_{\mathbf{x} \in \Omega} [\mathbf{x} - \mathbf{g}_\Omega][\mathbf{x} - \mathbf{g}_\Omega]^T dx$$

⁴Un contour proche est défini comme un contour dont les deux points extrémaux sont distants de moins de 10% de la longueur du contour.

où $S(\Omega)$ correspond à la surface de la région Ω et g_Ω au centre de gravité de Ω . Notons e_1 et e_2 les deux vecteurs propres de $Cov(\Omega)$ associés respectivement aux deux valeurs propres λ_1 et λ_2 .

Comme l'illustre la figure 2.4, l'ellipse d'inertie de Γ correspond à l'ellipse de centre g_Ω dont les axes sont orientés selon e_1 et e_2 et de dimensions respectives $\sqrt{\lambda_1}$ et $\sqrt{\lambda_2}$. Cette ellipse correspond donc à l'image du cercle unité par la transformation E :

$$\mathbf{E} = (Cov(\Omega)^{1/2} \mid g) \quad (2.1)$$

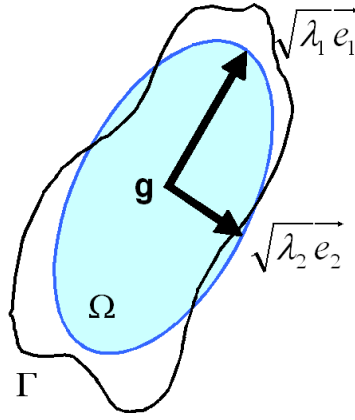


FIG. 2.4 – L'ellipse d'inertie d'un contour Γ .

Cette ellipse \mathcal{E} peut aussi être définie à l'aide de la distance de Malahanobis. On obtient alors :

$$\mathcal{E} = \{ x \in R^2 \text{ tel que } (x - g_\Omega)^T Cov(\Omega)^{-1} (x - g_\Omega) = 1 \}$$

Cette ellipse présente l'intérêt d'être covariante avec les transformations affines appliquées à Γ , tout comme son intérieur. Pour toute transformation affine A , on a $Cov(A\Omega) = A^T Cov(\Omega) A$ et $g_{A\Omega} = A g_\Omega$ (une démonstration de ces deux propositions est présentée dans [63]).

A partir de ces propositions, on peut démontrer que pour tout point x appartenant à l'intérieur de \mathcal{E} , Ax appartient à l'ellipse définie par la région $A\Omega$:

$$\begin{aligned} (Ax - g_{A\Omega})^T Cov(A\Omega)^{-1} (Ax - g_{A\Omega}) &= (x - g_\Omega)^T A^T (A Cov(\Omega) A^T)^{-1} A (x - g_\Omega) \\ &= (x - g_\Omega)^T A^T A^{-T} Cov(\Omega)^{-1} A^{-1} A (x - g_\Omega) \\ &= (x - g_\Omega)^T Cov(\Omega)^{-1} (x - g_\Omega) \\ &\leq 1 \end{aligned}$$

L'image de \mathcal{E} par la transformation A est donc incluse dans l'image de l'ellipse définie par $A\Omega$. D'une manière similaire, on peut démontrer que l'ellipse associée à $A\Omega$ est incluse dans \mathcal{E} .

L'ellipse d'inertie \mathcal{E} permet donc de définir une région covariante avec les transformations affines. Afin d'obtenir une région support couvrant une surface de l'objet suffisamment informative pour le calcul d'une signature, un facteur d'échelle $s = k.f$ est appliqué à \mathcal{E} . Le facteur k correspond à une constante fixée par l'utilisateur alors que f correspond la distance de Mahalanobis moyenne des points de contour de Γ au centre de la région support :

$$f = \frac{\sum_{x_i \in \Gamma} \sqrt{(x_i - g) \text{Cov}(\Omega)^{-1} (x_i - g)^T}}{\text{Card}(\Gamma)}$$

Ainsi, le paramètre k permet de définir la taille de la région support en fonction de la taille de Γ . L'influence de k sur la région support est illustrée dans la figure 2.5. Lorsque $k = 1$, la distance moyenne entre les points de contours de Γ et la frontière de la région support est nulle. Dans ce cas, la région support obtenue approxime Γ . Dans les différentes expérimentations présentées dans ce mémoire, ce facteur a été fixé à $k = 5$.

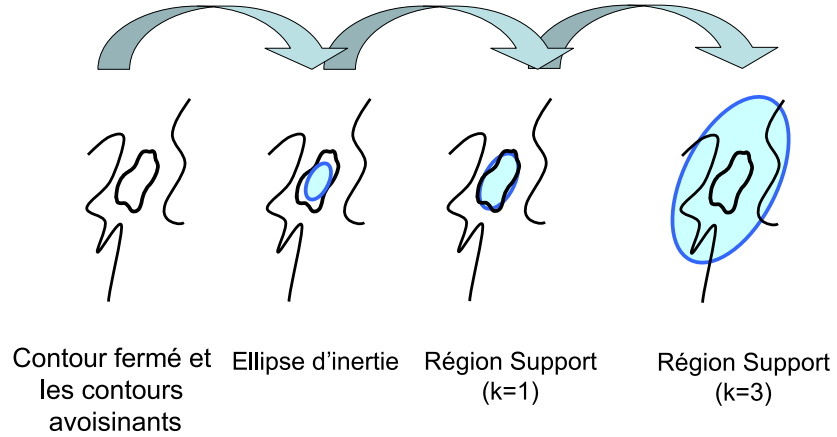


FIG. 2.5 – Calcul de la région support à partir d'un contour fermé.

2.1.2 Caractérisation de la région support

Chaque région support est caractérisée par deux signatures. La première est proche de la signature RIFT [44] alors que la seconde correspond à une signa-

ture *Shape Context* [6]. Afin que l'identité d'un descripteur soit invariante (ou du moins peu sensible) aux changements de points de vue, ces signatures ne sont pas calculées directement sur la région support mais sur la région support normalisée. Cette normalisation permet d'obtenir une unique représentation de la région quelque soit la transformation affine appliquée à celle-ci.

Normalisation de la région support

La normalisation d'une région support consiste donc à calculer une représentation de la région support qui soit invariante par toute transformation affine inversible. Puisque la région support correspond à une ellipse, la normalisation consiste simplement à redresser cette ellipse en un disque unitaire. Or, la région support correspond, à un facteur d'échelle s près, à l'ellipse d'inertie \mathcal{E} , cette dernière correspondant à l'image du cercle unité par la transformation \mathbf{E} (cf. équation 2.1). Par conséquent, l'a région support correspond à l'image du cercle unité par la transformation T :

$$\mathbf{T} = (s.Cov(\Omega)^{1/2} \mid \mathbf{g}) \quad (2.2)$$

la normalisation de cette région consiste simplement à lui appliquer la transformation affine \mathbf{F} :

$$\mathbf{F} = \mathbf{T}^{-1} \quad (2.3)$$

On désignera sous le terme *Région Support Normalisée* (ou RSN) l'image de la région support par cette transformation \mathbf{F} . La région support contenant des points de contours, la transformation \mathbf{F} est appliquée à la fois à la position et à l'orientation de chaque point de contour.

Cette normalisation présente néanmoins deux limitations. D'une part, le processus de normalisation ne contraint pas la rotation autour du centre de la RSN ; d'autre part la normalisation de contours discrets (ie série de pixels) entraîne l'apparition d'artefacts.

Problème de la rotation La transformation redressant une ellipse en un disque unitaire n'est pas unique. En effet, l'image du disque unitaire par une rotation est aussi un disque unitaire. Par conséquent, quelque soit la matrice de rotation R , l'image de la région support par $R.F$ correspond aussi une normalisation de celle-ci. Comme l'illustre la figure 2.6, la région normalisée est donc définie de manière unique modulo une rotation.

Afin de pouvoir supprimer ce dernier degré de liberté, certains travaux fixent une orientation à la RSN en recherchant une direction principale [51] ou un point

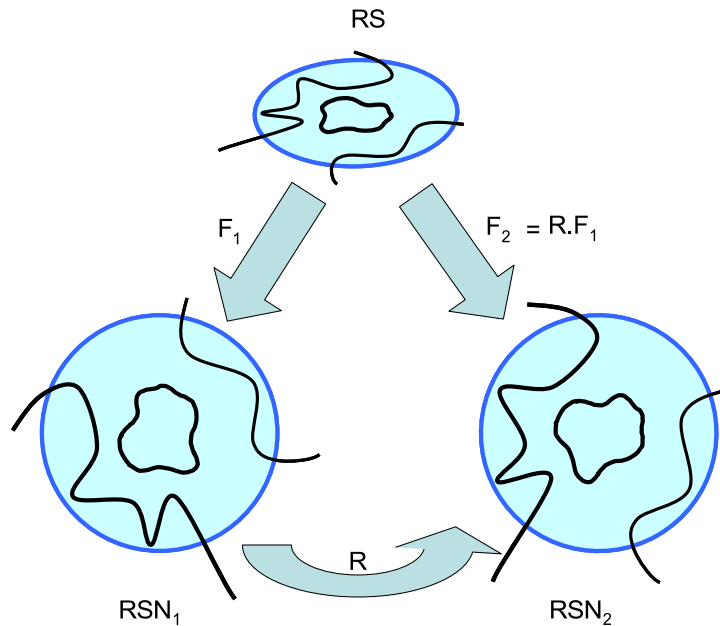


FIG. 2.6 – Normalisation de la région support à une rotation près. *La région support (RS) peut être normalisée selon plusieurs transformations (F_i). Les régions support normalisées résultantes (RSN_i) sont reliées par des rotations (R_i).*

extrémal [64] du contour fermé. Ces solutions sont généralement peu stables pour les pièces mécaniques. En effet, ces pièces présentent généralement de nombreux alésages de forme quasi-circulaire. Ces derniers ne permettent donc pas de définir de manière stable un point extrémal et ou une direction principale.

Dans le cadre des travaux présentés, deux solutions ont été envisagées. La première consiste à gérer ce dernier degré de liberté au niveau de la signature (usage d'une signature invariante aux rotations). La seconde solution consiste à gérer les variations de signature induites par la rotation lors de la comparaison de deux signatures. Ces deux solutions sont présentées dans les paragraphes suivants.

Artefacts liés à la normalisation Si la normalisation est parfaitement définie pour des contours continus, son application à des contours discrets induit des artefacts : les contours de la RSN sont segmentés (cf. fig. 2.7). Cette segmentation est d'autant plus importante que l'excentricité de l'ellipse définissant la région support est élevée. Sans compensation de ces artefacts, la RSN ne peut être considérée comme invariante aux transformations affines.

Afin de réduire l'influence de ces artefacts, une pondération est attribuée à

chaque pixel de contour de la RSN. La fragmentation d'un contour dépendant de l'orientation de celui-ci par rapport au grand axe de la région support (la fragmentation est maximale lorsque le contour est orthogonal à la direction du grand axe), nous avons choisi de fixer la pondération en fonction de cette orientation. Pour estimer ce poids, on attribue à chaque pixel de contour de la RS un segment unitaire tangent au contour. A l'issue de la normalisation, ce segment initialement unitaire possède une longueur t . On attribue alors au pixel de la RSN le poids t (cf. figure 2.8). La figure 2.9 illustre ce procédé et montre comment cette pondération compense les effets de la segmentation.

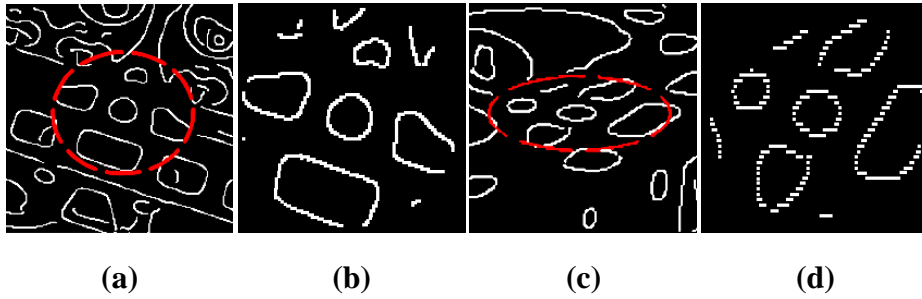


FIG. 2.7 – Artefacts de segmentation lors de la normalisation d'une région support. (a) et (c) : les régions supports. (b) et (d) : les régions supports (a) et (c) après normalisation. On note un effet de segmentation sur (d) alors que celui-ci est invisible sur (b).

Signature invariante aux rotations : STHGI

Principe La signature STHGI caractérise la distribution des orientations des points de contour d'une RSN en fonction de la distance au centre de cette RSN. Nous appelons orientation d'un point de contour l'orientation du vecteur gradient en ce point.

La figure 2.10 illustre la méthode de calcul de cette signature. La RSN est découpée en N anneaux concentriques et la distribution des orientations des contours est calculée dans chacun de ces anneaux. Le calcul de cette distribution intègre le poids associé à chaque point de contour lors de l'étape de normalisation, ceci afin de compenser l'effet de discrétisation et d'obtenir une signature invariante au facteur d'échelle. Cette distribution est représentée sous la forme d'un histogramme des orientations des contours de taille M , l'espace des valeurs angulaires ayant été discrétisé en M secteurs angulaires (typiquement $M = 4$).

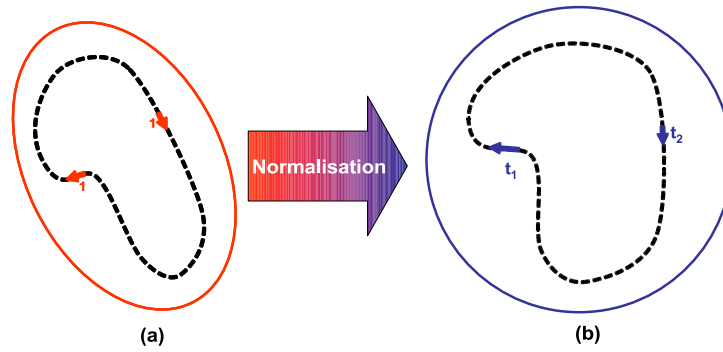


FIG. 2.8 – Calcul du poids d'un point de contour. *En rouge : l'ellipse représente une région support d'un contour fermé (en pointillé). Le poids (unitaire) et l'orientation de deux points de contours sont représentés par l'orientation et la longueur de deux flèches rouges. L'ellipse bleue représente cette même région support après normalisation. En raison de leur orientation différentes, les deux points de contours se voient attribuer des poids différents à l'issue de la normalisation ($t_1 > t_2$). Le poids de chacun des points de contours correspond à la longueur de la flèche qui lui est attribuée.*

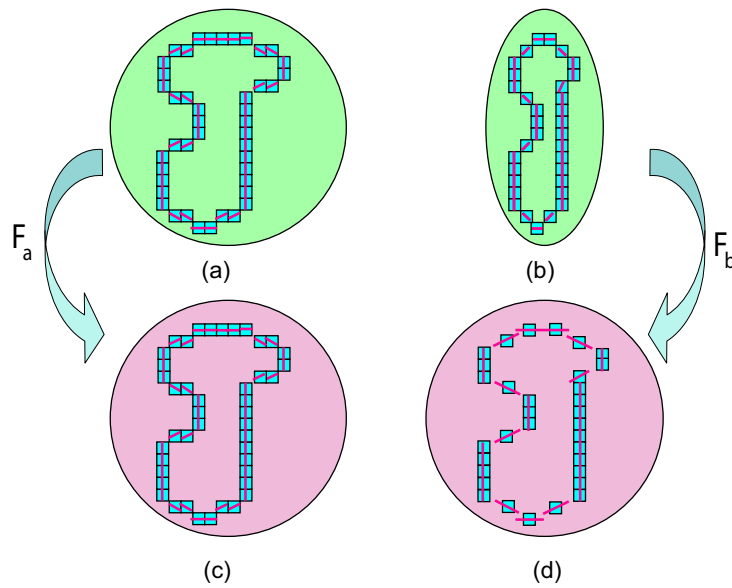


FIG. 2.9 – Pondération des points de contours : (a) et (b) : la même région support observée selon deux points de vue différents. (c) et (d) : les régions supports normalisées de (a) et (b). La longueur d'un segment rouge correspond au poids du pixel de contour. La pondération compense l'effet de discrétisation et rend les RSN équivalentes.

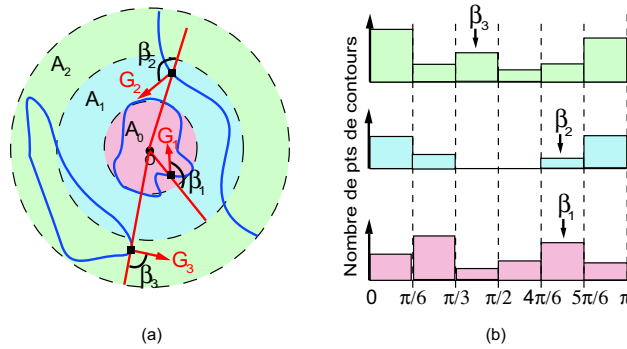


FIG. 2.10 – Calcul de la signature STHGI. A gauche, la région support normalisée est découpée en 3 anneaux (A_i). L'orientation d'un point de contour (carré noir) correspond à l'angle β_i entre la normale au contour (G_i) et l'axe reliant le point de contour à O (trait rouge). A droite, la signature STHGI est représentée sous la forme de 3 histogrammes des orientations des points de contours, chaque histogramme caractérisant l'anneau de couleur correspondante.

Afin de rendre la signature invariante par rotation, l'orientation d'un point de contour est exprimée par rapport à l'axe reliant le centre de la région au point de contour, ce qui correspond aux angles β_i de la figure 2.10. L'orientation est ainsi estimée dans un repère local qui est indépendant de l'orientation de la RSN. La distance au centre de la RSN étant elle aussi indépendante de l'orientation de la RSN, la signature STHGI est donc bien invariante aux rotations.

Enfin, on notera que l'espace angulaire se limite à l'intervalle $[0, \pi[$. En effet, seule la direction du contour est prise en compte lors du calcul de l'histogramme. Le sens est ignoré car le vecteur gradient peut s'inverser en fonction de l'éclairage⁵.

La signature STHGI correspond donc à la concaténation de N histogrammes de taille M , soit un histogramme global de taille $M \times N$.

Avantages/Inconvénients Le principal intérêt de cette signature réside dans son invariance aux rotations. Grâce à cette propriété et à la normalisation, une unique signature permet de caractériser une région support, et cela quelque soit les transformations affines appliquées à cette région. Par conséquent, il est aisé de comparer les signatures de deux régions supports. La distance du χ^2 , qui présente

⁵Lorsqu'il s'agit de surfaces métalliques, une discontinuité clair/sombre peut devenir sombre/clair à la suite d'une variation des conditions d'illumination.

l'avantage d'être rapide à évaluer, peut être employée pour comparer les signatures STHGI. Le coût du calcul de la signature STHGI étant faible lui aussi, elle peut être employée pour comparer rapidement un grand nombre de descripteurs.

L'inconvénient majeur de cette signature vient de la faible caractérisation spatiale de la distribution des points de contours. En effet, la répartition des orientations des contours dans un anneau ne caractérise pas la répartition spatiale des points de contours dans l'anneau. Par conséquent, des régions supports très différentes peuvent présenter une signature similaire comme l'illustre la figure 2.11.

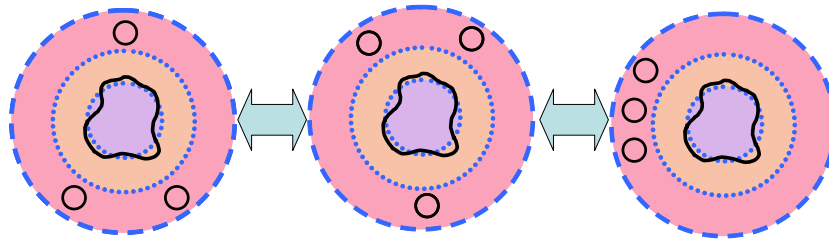


FIG. 2.11 – Exemples de RSN partageant la même signature. *En ne caractérisant pas la répartition spatiale au sein des anneaux, le STHGI ne permet pas de différencier ces trois RSN.*

La signature *Shape Context*

Principe La seconde signature utilisée pour caractériser une RSN correspond à la signature de *Shape Context* [6]. Il s'agit d'une signature caractérisant la distribution des points de contours. Cette caractérisation est obtenue en découpant la région support avec une grille radiale⁶ de N anneaux et Q secteurs (typiquement $N = 3$ et $Q = 64$) telle qu'illustrée dans la figure 2.12. Cette grille radiale décompose donc la région support en $N \times Q$ secteurs. Pour chacun des secteurs, la somme des poids des points de contour dans celui-ci est estimée. Les résultats obtenus pour les différents secteurs sont stockés sous la forme d'une matrice $N \times Q$. Cette matrice correspond à la signature du *Shape Context*.

On notera que le découpage radial de la RSN fixe une orientation : l'origine du découpage en secteurs. La signature étant dépendante du choix de cette origine,

⁶Généralement, la grille utilisée est uniforme dans un espace log-polaire, ceci afin d'avoir une signature d'autant plus sensible à la position des points de contour que ceux-ci sont proches du centre de la région support.

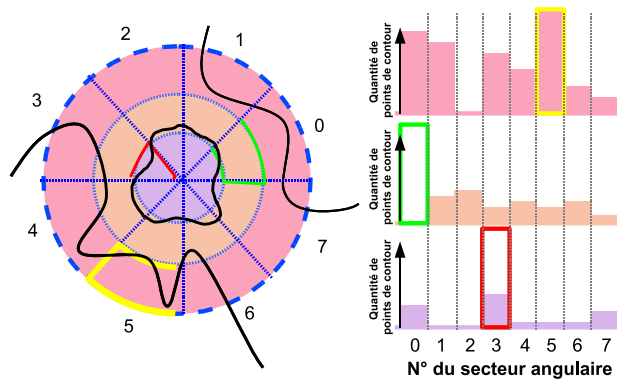


FIG. 2.12 – Calcul de la signature *Shape Context* : à gauche, la région support normalisée découpée par une grille radiale de 3 anneaux et 8 secteurs ; à droite, la signature représentée sous la forme de 3 histogrammes. Chaque histogramme caractérise l'anneau de couleur correspondante de la grille radiale. Chaque barre de l'histogramme encadrée d'une couleur correspond à la case encadrée de la même couleur dans la grille radiale. La hauteur d'une barre de l'histogramme correspond au nombre de points de contour contenus dans la case de la grille correspondante.

le Shape Context n'est pas une signature invariante aux rotations. Les effets d'une rotation peuvent d'ailleurs être décrits selon deux cas de figure :

1. la rotation est inférieure à $2\pi/Q$, alors la signature subit des changements difficilement prévisibles.
2. la rotation vaut $2k\pi/Q$ avec $k \in \mathbb{N}$, alors la signature subit k décalages circulaires de ses colonnes.

Pour une rotation quelconque, les effets correspondent à une combinaison des deux cas précédemment exposés.

Si les effets liés au premier cas sont difficiles à gérer, on constate expérimentalement que ceux-ci deviennent négligeables lorsque Q devient grand (e.g. $Q > 30$). Dans ce cas, les effets liés au second cas étant connus, il est possible d'intégrer la gestion de ce phénomène au niveau de la distance utilisée pour comparer deux signatures.

La distance retenue repose sur une corrélation normée centrée (ZNCC). Afin de prendre en compte les permutations circulaires pouvant être engendrées par une rotation de la RSN, cette distance évalue une série de Q corrélations avec un décalage circulaire d'une colonne à chaque application successive. La distance

est alors évaluée à partir du score de corrélation maximal obtenu au cours de ces Q applications. Pour définir cette distance, notée D_{SC} , entre deux signatures de *Shape Context*, notons :

- s_j : la signature *Shape Context* de la région j avec Q secteurs angulaires,
- s_i : la signature *Shape Context* de la région i avec Q secteurs angulaires,
- s_i^n : la signature s_i après n décalages circulaires de ses colonnes, avec $0 \leq n < Q$,
- $zncc(s_i, s_j)$: le score de corrélation normée centrée entre s_i et s_j .

La distance $D_{SC}(s_i, s_j)$ est alors définie par :

$$D_{SC}(s_i, s_j) = 1 - \max_{n=1..Q} (zncc(s_i^n, s_j)) \quad (2.4)$$

Avantages/Inconvénients Cette signature présente l'avantage d'être plus discriminante que la signature STHGI, ceci du fait de la meilleure caractérisation spatiale de la RSN. Cependant, cette caractérisation spatiale ne permet pas d'obtenir une signature invariante aux rotations. Par conséquent, une distance coûteuse en temps de calcul doit être employée pour comparer deux signatures. Cette signature n'est donc pas adaptée à la recherche de signatures similaires dans une grande base de descripteurs. La signature *Shape Context* ne vient donc pas remplacer la signature STHGI mais, au contraire, vient la compléter. Nous verrons dans le chapitre 3 que la complémentarité de ces deux signatures permet d'établir un processus de mise en correspondance à la fois rapide et discriminant.

2.2 Organisation des primitives visuelles en modèle géométrique 3D

La section précédente a présentés les primitives visuelles du modèle mixte permettant de caractériser l'apparence de l'objet. Pour obtenir un modèle mixte, ces primitives doivent être associées à des primitives 3D et organisées spatialement pour de former un modèle géométrique 3D, comme l'illustre la figure 2.13.

Dans le cas des descripteurs locaux de contours, la primitive 3D retenue correspond à un point 3D. Plus précisément, ce point 3D correspond au point de l'objet se projetant sur le centre g de la région support du descripteur. Puisque l'image dont est extrait ce descripteur est recalée par rapport au modèle 3D de l'objet, la position 3D de ce point peut être déterminée. Les vues à partir desquelles sont extraits les descripteurs locaux étant parfaitement calibrées (les paramètres intrinsèques et extrinsèques de la caméra sont connus pour chaque vue), un lancer de

2.2. ORGANISATION DES PRIMITIVES VISUELLES EN MODÈLE GÉOMÉTRIQUE 3D 47

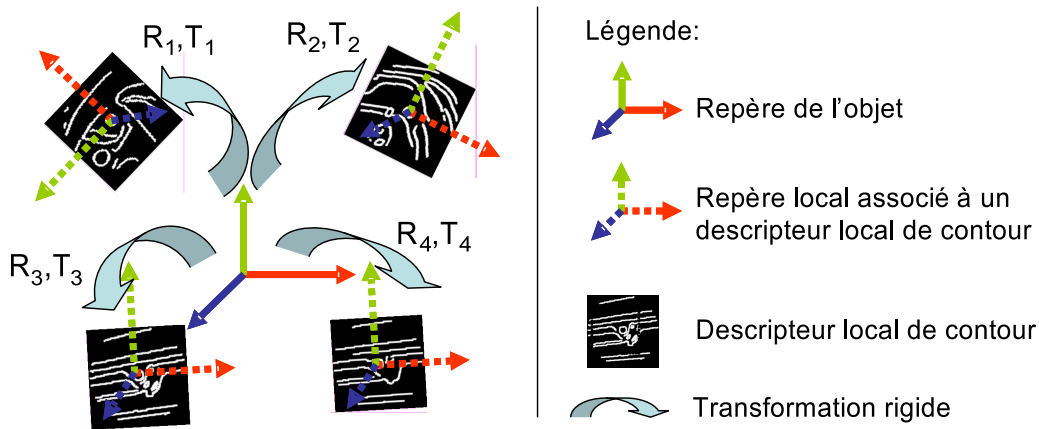


FIG. 2.13 – Le modèle géométrique 3D *Le modèle géométrique associée à chaque descripteur local de contour un repère 3D, définissant la position et l'orientation 3D de la primitive visuelle, dans le repère global de l'objet.*

rayon permet de déterminer, pour tout point P de l'image, le point 3D de la surface de l'objet se projetant en P . Pour cela, un rayon, passant par le centre optique de la caméra et par le point P est lancé. Le premier point d'intersection entre ce rayon avec le modèle 3D correspond au point 3D de l'objet se projetant sur P . Si le rayon n'intersecte pas le modèle CAO, cela signifie que P n'appartient pas à l'image de l'objet.

Il est donc possible d'estimer la position 3D du centre de la région support d'un descripteur. Cependant, rien n'assure que le centre de la région support appartienne à l'objet. En effet, puisqu'une grande partie des descripteurs caractérise des alésages, il est fort probable que le centre de la région support corresponde au centre d'un trou dans l'objet. D'autre part, la surface de l'objet correspondant à la région support n'est généralement nécessairement plane. La position 3D du point de l'objet se projetant au centre de la région support peut varier en fonction du point de vue depuis lequel l'objet est observé. Or, il est nécessaire que cette position soit stable pour que cette primitive puisse être utilisée dans un calcul de pose.

Une solution permettant de résoudre ces deux problèmes consiste à préalablement approximer la surface de l'objet correspondant à la région support par un plan 3D noté Π . Ceci permet de supprimer le problème des trous mais aussi de se placer sous l'hypothèse d'un objet localement plan. La région support étant définie à partir de la position 2D des points du contour fermé, le plan 3D est estimé à

partir de la position 3D de ces même points estimés de la manière exposée préalablement. Plus précisément, Π correspond au plan 3D qui minimise la distance orthogonale entre les points 3D correspondant au contour fermé et le plan Π . La position 3D du centre g de la région support est alors calculée en utilisant le plan Π au lieu du modèle CAO. Notons G cette position 3D.

A partir du plan Π et du point G il est possible d'associer un repère local à chaque descripteur. Le plan Π définit le plan xy du repère alors que le point G en définit l'origine. L'axe z du repère correspond donc à la normale du plan Π , alors que les axes x et y sont fixés arbitrairement dans le plan Π . Ce repère ajoute à chaque descripteur une orientation 3D en plus la position 3D définie par G . Il est alors possible de calculer la transformation rigide reliant le repère de l'objet et le repère associé au descripteur (les transformations R_i, T_i de la figure 2.13). Ces changements de repères constituent les contraintes géométriques 3D du modèle de l'objet. Le fait d'exprimer ces contraintes en terme de changements de repères, et non pas de simple, translation 3D, sera utile lors de l'étape de filtrage géométrique des appariements présenté dans la section 4.1.

2.3 Extension du modèle : le motif 2D/3D

Dans le chapitre 4, un processus permettant de filtrer les faux appariements sera proposé. Ce processus nécessite d'établir un grand nombre d'hypothèses de poses à partir d'appariements entre les descripteurs locaux issus de l'image et ceux du modèle. Pour obtenir un processus rapide, il est souhaitable de pouvoir estimer une pose à partir d'un nombre minimal d'appariements. L'objectif de cette section est de présenter une extension du modèle permettant de calculer une pose à partir d'un seul appariement.

Le modèle géométrique, tel qu'il vient d'être présenté, n'associe qu'un seul point 3D à chaque descripteur du modèle. Or, un calcul de pose nécessite au minimum 3 appariements entre des points 2D de l'image et des points 3D du modèle. Pour qu'une pose puisse être estimée à partir d'un seul appariement avec une primitive du modèle, nous avons choisi d'enrichir l'information 3D contenue dans celle-ci. La solution proposée dans cette section consiste à associer un **motif 2D/3D** à chaque primitive du modèle.

Le motif 2D/3D correspond à une description conjointe de l'apparence 2D d'une région support et de la géométrie 3D qui lui est sous-jacente. La description de l'apparence correspond à l'image de contour du descripteur. Pour que cette

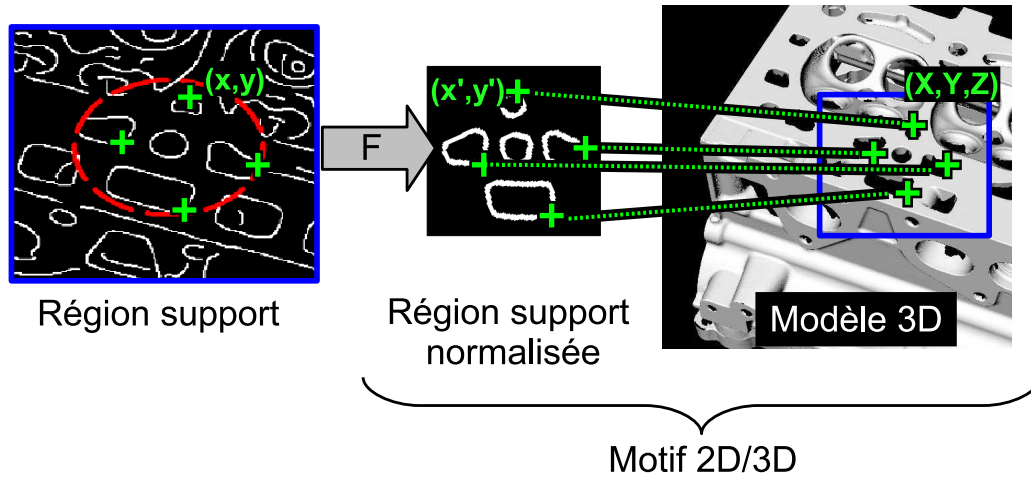


FIG. 2.14 – Principe du motif 2D/3D. Chaque point (x',y') de la région support normalisée correspond à un point (x,y) dans le repère de l'image (i.e. région support avant la normalisation). L'image d'apprentissage étant recalée avec le modèle 3D, on connaît la position 3D (X,Y,Z) du point de l'objet qui se projette en (x,y) . Le motif 2D/3D correspond à l'ensemble des couples $((x',y'), (X,Y,Z))$.

représentation soit intrinsèque au descripteur (i.e. robuste aux changements de point de vue), la position (x,y) d'un point de contour est exprimée dans le repère de la région support normalisée. Notons (x',y') la position de ce même point de contour après normalisation (cf. fig. 2.14).

A cette description 2D de la région support est associée une description 3D. A chaque point de contour (x',y') est associée la position 3D (X,Y,Z) de ce même point. Elle correspond au point 3D de l'objet se projetant en (x,y) (pour le calcul de la position 3D d'un point de contour, se référer à la section 2.2).

Le motif 2D/3D d'un descripteur M correspond alors à l'ensemble des couples $\{(x',y'), (X,Y,Z)\}$ obtenus à partir de la RSN du descripteur M (cf. figure 2.14). Puisque les coordonnées 2D sont exprimées dans le repère normalisé de la région support, ceux-ci sont invariants aux changements de point de vue d'une caméra affine. De même, les coordonnées 3D étant exprimées dans le repère de l'objet, celles-ci sont invariantes aux changements de point de vue. Par conséquent, le motif 2D/3D peut être considéré comme une description intrinsèque de la région support.

L'exploitation du motif 2D/3D pour le calcul d'une pose à partir d'un seul appariement sera exposée dans le chapitre 4.

2.4 Conclusion

Dans ce chapitre, un modèle mixte combinant une description de l'apparence et de la géométrie de l'objet a été proposé. Ce modèle peut être décrit comme un modèle géométrique 3D dont les primitives sont des descripteurs locaux. L'application visée étant le recalage de pièces mécaniques comportant des alésages, les descripteurs locaux employés sont des descripteurs locaux de contour caractérisant le voisinage d'un contour fermé. Une position et une orientation 3D sont attribuées à chaque descripteur local en étudiant la position 3D des points de surface de l'objet générant les contours à l'image.

De plus, afin de permettre l'estimation de la pose de l'objet à partir d'un seul appariement entre un descripteur du modèle et un descripteur local de l'image, chaque descripteur du modèle est associé à une description de cette surface 3D qui lui est sous-jacente : le motif 2D/3D.

La construction de ce modèle est réalisée hors ligne par un algorithme automatique utilisant un ensemble fini de vues recalées de l'objet et son modèle CAO.

Le modèle de l'objet étant défini, celui-ci peut être exploité pour identifier les primitives visuelles détectées dans l'image et ensuite en déduire la pose de l'objet.

Chapitre 3

Mise en correspondance selon l'apparence

Objectif : Pour pouvoir déterminer la pose d'un objet à partir d'une image, il est nécessaire d'identifier les manifestations visuelles de ce dernier dans l'image. Cette identification consiste à établir, pour chacune de ces primitives, une correspondance avec une primitive du modèle. L'objectif de ce chapitre est de proposer un processus permettant de réaliser cette tâche à l'aide du modèle présenté au chapitre précédent.

Etablir une mise en correspondance, entre la scène observée et le modèle de l'objet, nécessite d'identifier les manifestations visuelles correspondant à l'objet. Comme il a été présenté dans l'état de l'art, les descripteurs locaux possèdent généralement un pouvoir discriminant suffisant pour pouvoir les identifier individuellement. Cette propriété permet d'apparier les primitives visuelles indépendamment les unes des autres, ce qui permet d'obtenir un processus de faible complexité.

Cependant, deux difficultés se posent avec ce type de solution. D'une part, la signature associée à une région support n'est généralement pas totalement invariante aux conditions d'observations (point de vue, illumination de la scène, bruit dans l'image,...). Cette signature présente donc une certaine variabilité qui doit être prise en compte, lors de la mise en correspondance, afin de pouvoir décider si un appariement est correct ou non. Cependant, cette prise en compte tend à réduire le pouvoir de discriminant de la signature. Le critère de mise en correspondance est alors nécessairement un compromis entre l'acceptabilité d'une certaine variabilité et le pouvoir discriminant de la signature.

D'autre part, un objet peut comporter de nombreux motifs répétitifs, c'est à

dire différentes régions de l'objet possédant le même aspect. Comme l'illustre la figure 3.1, ce cas de figure est particulièrement fréquent chez les pièces mécaniques, celles-ci comportant généralement des structures géométriques se répétant à leur surface (e.g. série d'alésage, etc.). Ces motifs répétitifs induisent, au niveau du modèle de l'objet, la présence de descripteurs partageant la même signature mais pas la même position. Ces primitives du modèle ne pouvant être distinguées par leur signature, celles-ci peuvent engendrer un nombre important de faux appariements lors de l'étape de mise en correspondance.

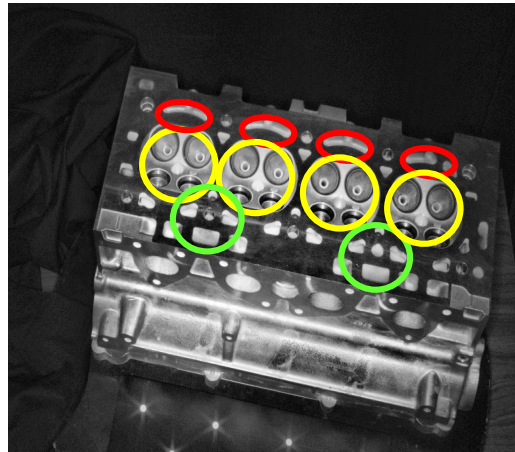


FIG. 3.1 – Exemples de motifs répétitifs sur une culasse de moteur.

Ce chapitre propose un processus de mise en correspondance s'appuyant sur le modèle d'apparence sous-jacent au modèle de l'objet. Ce processus vise notamment à résoudre les deux difficultés présentées ci-dessus. La section 3.1 explique comment les deux signatures, qui caractérisent chacun des descripteurs du modèle, sont exploitées pour établir une mise en correspondance à la fois rapide et discriminante. La section 3.2 présente une solution permettant de décider de la validité d'un appariement, celle-ci autorisant une certaine variabilité de la signature du descripteur image vis à vis de celle à laquelle elle est appariée tout en assurant un pouvoir discriminant élevé. La section 3.3 propose une adaptation du modèle de l'objet et du processus de mise en correspondance afin d'intégrer la notion de motifs répétitifs. Enfin, ce chapitre se termine avec une évaluation du processus de mise en correspondance (section 3.4).

3.1 Processus de mise en correspondance

Avant de présenter le processus de mise en correspondance, il est nécessaire de définir quelques notations. Notons $\Lambda = \{I_i, i = 1..N\}$ l'ensemble des descripteurs extraits de l'image. Notons $\Omega = \{M_j, j = 1..K\}$ l'ensemble des descripteurs formant le modèle de l'objet. D'autre part, on désignera par $\langle I_i, M_j \rangle$ la mise en correspondance entre le descripteur image I_i et le descripteur du modèle M_j .

Initialement, chaque descripteur de l'image I peut potentiellement être apparié à l'un des K descripteurs du modèle. Il y a donc K hypothèses d'appariement à envisager pour chaque I_i .

Pour déterminer la primitive du modèle dont la signature est la plus proche du descripteur I_i , il est nécessaire d'évaluer K distances. Pour appairer l'ensemble des éléments de Λ , $K.N$ distances doivent donc être évaluées. Si la signature de *Shape Context* offre le meilleur pouvoir discriminant et correspond donc au meilleur critère de sélection, la distance qui lui est associée présente un coût calculatoire trop élevé pour envisager KN évaluations. S'agissant de la signature STHGI, celle-ci est associée à une distance rapide à calculer mais offre un pouvoir discriminant insuffisant pour assurer à elle seule la réussite du processus de mise en correspondance.

Afin d'obtenir un processus de mise en correspondance à la fois rapide et discriminant, la solution proposée se déroule en deux passes au cours desquelles les deux signatures sont exploitées.

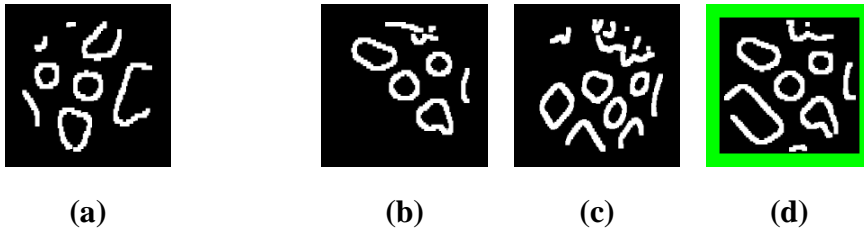


FIG. 3.2 – Mise en correspondance en deux passes. (a) correspond à la région support normalisée issue de l'image. (b), (c) et (d) sont les régions supports normalisées du modèle sélectionnées après la première passe (ici $T=3$). La région support normalisée encadrée en vert correspond à celle sélectionnée lors de la seconde passe et appariée à (a).

La première étape consiste à réduire, pour chaque I_i , l'ensemble des hypothèses d'appariements envisageables à un sous-ensemble de taille T fixée (typiquement $T = 5$). Il s'agit donc de sélectionner, parmi les K hypothèses possibles,

les T hypothèses les plus probables. La signature STHGI est employée pour cette tâche. La rapidité du calcul de la distance permet d'explorer rapidement un grand nombre d'hypothèses alors que son pouvoir discriminant moyen est compensé par la sélection des T meilleures hypothèses au lieu de la seule meilleure.

La seconde étape consiste à identifier lequel des T descripteurs restants correspond au descripteur I_i . En fixant $T \ll K$, le nombre de distances évaluées est en $o(N)$. Cette réduction de la complexité permet alors d'employer la signature de *Shape Context*. On note $\langle I_j, M_{j(i)} \rangle$ l'appariement issu de cette dernière étape.

L'usage combiné des deux signatures permet d'obtenir un processus de mise en correspondance discriminant sans pour autant avoir des temps de calcul rédhibitoires. Cependant, pour achever ce processus, deux points restent à étudier. Tout d'abord, il s'agit de définir un seuil sur la distance entre les signatures afin de déterminer si un appariement est correct ou non. Ensuite, il s'agit d'intégrer la notion de motifs répétitifs au processus de mise en correspondance afin de ne pas manquer certains appariements corrects. Ces deux points sont étudiés dans les sections suivantes.

3.2 Modélisation de la variance de la signature d'un descripteur du modèle

On qualifiera de *descripteurs compatibles*, deux descripteurs dont les signatures *Shape Context* sont suffisamment similaires pour que ceux-ci caractérisent la même région de l'objet. Pour décider si deux descripteurs sont compatibles, il est nécessaire de fixer un seuil ς sur la distance D_{SC} (cf. équation 2.4) qui les sépare. Ce seuil doit permettre l'identification d'un descripteur en dépit des erreurs qui peuvent affecter le calcul de sa signature. Cependant, ce seuil doit aussi assurer un pouvoir discriminant minimal : la probabilité que deux signatures tirées au hasard aient une distance inférieure à ς doit être inférieure à un seuil fixé.

Celui-ci étant difficile à définir de manière analytique, il est fixé empiriquement. La solution proposée consiste à étudier la distribution des distances entre descripteurs pour le déterminer.

Une première solution consiste à réaliser un grand nombre d'expériences aléatoires, chaque expérience consistant à tirer deux signatures au hasard et à estimer la distance D_{SC} qui les sépare. Comme l'illustre la figure 3.3, les résultats laissent apparaître une distribution de distances pouvant être approximées par une distribution gaussienne $\mathcal{N}(\mu, \sigma)$. Par conséquent, le seuil est fixé à $\varsigma = \mu - 2\sigma$ afin d'assurer que la probabilité d'apparier deux descripteurs générés aléatoirement soit de l'ordre de 2,5%.

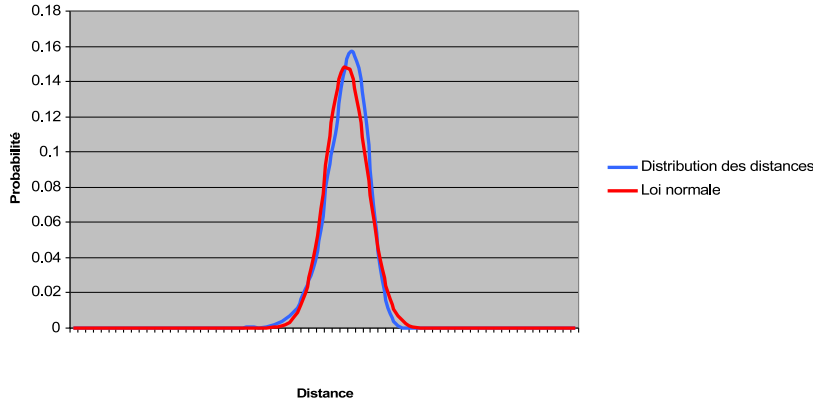


FIG. 3.3 – Approximation de la distribution des distances par une gaussienne. La courbe bleue représente la distribution des distances obtenues lors des expériences pour une signature de Shape Context de 3 anneaux et 64 secteurs. La distance moyenne est $\mu = 0.41$ et $\sigma = 0.0016$. En rouge, la distribution selon la loi normale définie par μ et σ .

Cependant, la distribution des distances a été obtenue sous l'hypothèse qu'une signature corresponde à un ensemble de variables aléatoires indépendantes. Or, cette hypothèse implique que les points de contours soient répartis de manière aléatoire dans la RSN, ce qui est bien évidemment faux. Le seuil ainsi estimé n'est donc pas pertinent pour une application réelle.

Pour remédier à ce problème, le seuil est estimé à partir d'une distribution de distances obtenue à partir de données réelles. Plus précisément, la base d'apprentissage employée correspond à l'ensemble des descripteurs du modèle. Ainsi, on calcule, pour chaque descripteur $M_j \in \Omega$, la distribution des distances entre la signature de ce descripteur M_j et le reste des signatures du modèle. Expérimentalement, on constate que l'approximation de cette distribution par une gaussienne reste valable (figure 3.4). Néanmoins, la présence de valeurs aberrantes dans cette distribution rend nécessaire l'usage d'outils de statistique robuste lors du calcul du seuil de distance. Ainsi, la distance moyenne μ est remplacée par la médiane et l'écart type σ est remplacé par la médiane des déviations absolues (Median Absolute Deviation ou MAD).

Expérimentalement, on constate des variations notables sur la valeur du seuil en fonction du descripteur M_j considéré. Ces variations reflètent la différence de pouvoir discriminant entre les différents descripteurs. Par conséquent, chaque descripteur M_j du modèle se voit attribuer un seuil spécifique calculé à partir de la seule distribution des distances des signatures du modèle à M_j . La figure 3.5 repré-

sente la distribution des distances pour un descripteur particulier. Les descripteurs du modèle dont la distance est inférieure au seuil de distance sont représentés par leur RSN. On peut constater que ces descripteurs partagent une apparence similaire. Ceux-ci correspondent soit à une même région de l'objet observée à partir d'un autre point de vue¹, soit à une autre région de l'objet ayant la même apparence. L'identification de ces descripteurs particuliers est utilisée, dans la section suivante, pour déterminer groupes de motifs répétitifs.

3.3 Gestion des motifs répétitifs

On désigne par "motifs répétitifs", un ensemble de régions de l'objet partageant la même apparence mais correspondant à des sites différents. Il s'agit d'un cas de figure fréquent avec les pièces mécaniques puisque celles-ci comportent généralement des structures géométriques (eg. une série de trous) répétées à la surface de l'objet (figure 3.1). Ces régions partageant la même signature, un processus de mise en correspondance basé sur l'apparence ne peut différencier les différentes instances d'un même motif répétitif. Ces motifs sont donc à la source de nombreuses erreurs d'appariements. De plus, puisque le modèle de l'objet est établi à partir de plusieurs vues, une instance d'un motif répétitif génère autant de descripteurs dans le modèle que le nombre de vues dans lesquelles elle est visible. Puisque le modèle compte généralement plusieurs descripteurs pour chaque instance d'un motif répétitif, le nombre de descripteurs du modèle partageant une même signature peut donc être particulièrement élevé. Afin d'illustrer ce problème, les trois images d'apprentissage utilisées pour former le modèle de la culasse dans les expériences des chapitres 5 et 6 sont présentées dans les figures 3.6, 3.8 et 3.10. Les ensembles de régions supports² extraites de chacune d'entre elles sont respectivement illustrés dans les figures 3.7, 3.9 et 3.11. A titre d'exemples, les descripteurs de quatre groupes de motifs répétitifs sont indiqués à la fois dans les images d'apprentissages et parmi les régions supports normalisées. On peut ainsi remarquer que le modèle ainsi établi comporte un grand nombre de régions supports partageant une apparence très similaire. Ce nombre élevé est dû au grand nombre de motifs répétitifs sur ce type de pièce ainsi qu'à leur observation dans les différentes vues d'apprentissages.

¹Le modèle étant élaboré à partir de plusieurs vues, une même région de l'objet peut engendrer autant de descripteurs dans le modèle que le nombre de vues dans lesquelles elle est visible.

²Les régions supports présentées sont issues d'une simple extraction de contours. Dans la pratique, les régions supports sont extraites à partir des contours issus d'une extraction multi-résolution.

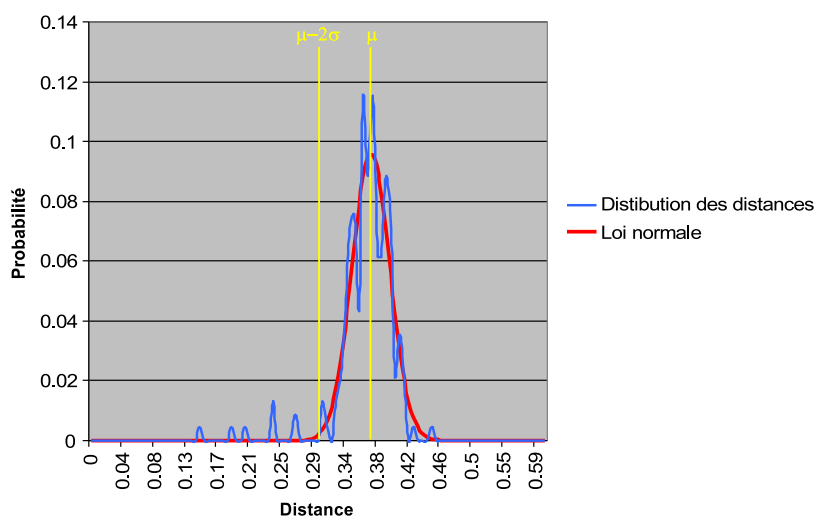


FIG. 3.4 – Distribution des distances au descripteur 37. La courbe bleue représente la distribution des distances au descripteur 37. En rouge, la distribution selon la loi normale calculée à partir de la médiane et du MAD de la distribution des distances.

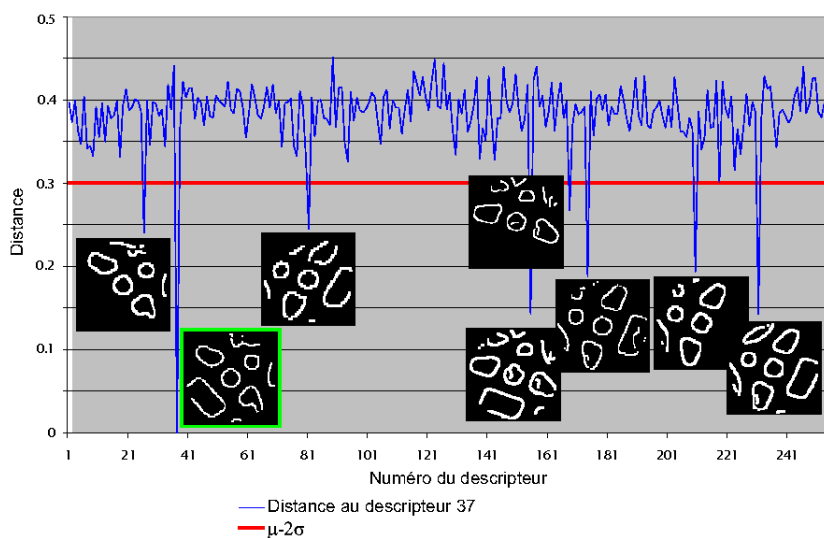


FIG. 3.5 – Exemple de distribution des distances pour une signature *Shape Context*. La courbe bleue correspond à la distribution des distances au descripteur numéro 37. La droite rouge correspond au seuil du descripteur 37. Les imageriees représentent les RSN des descripteurs dont la distance est inférieure au seuil.

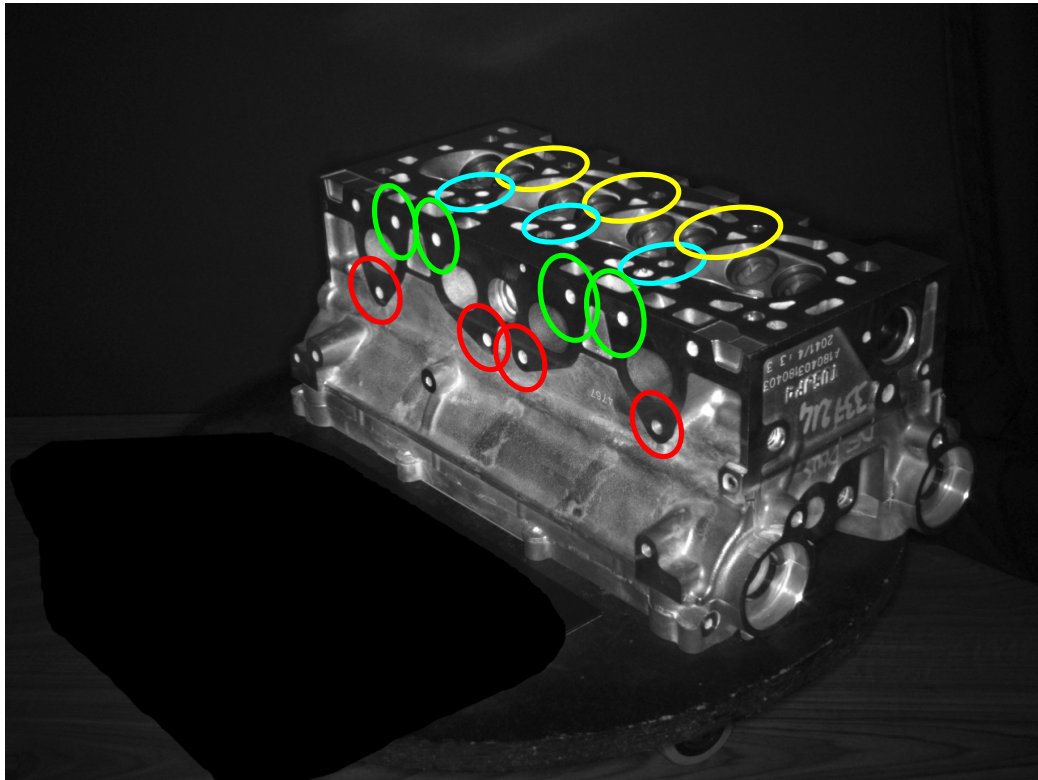


FIG. 3.6 – Première image d'apprentissage de la culasse. *Les ellipses de même couleur correspondent aux différentes régions supports d'un même motif répétitif. Les régions supports normalisées correspondantes peuvent être observées dans la figure 3.7.*

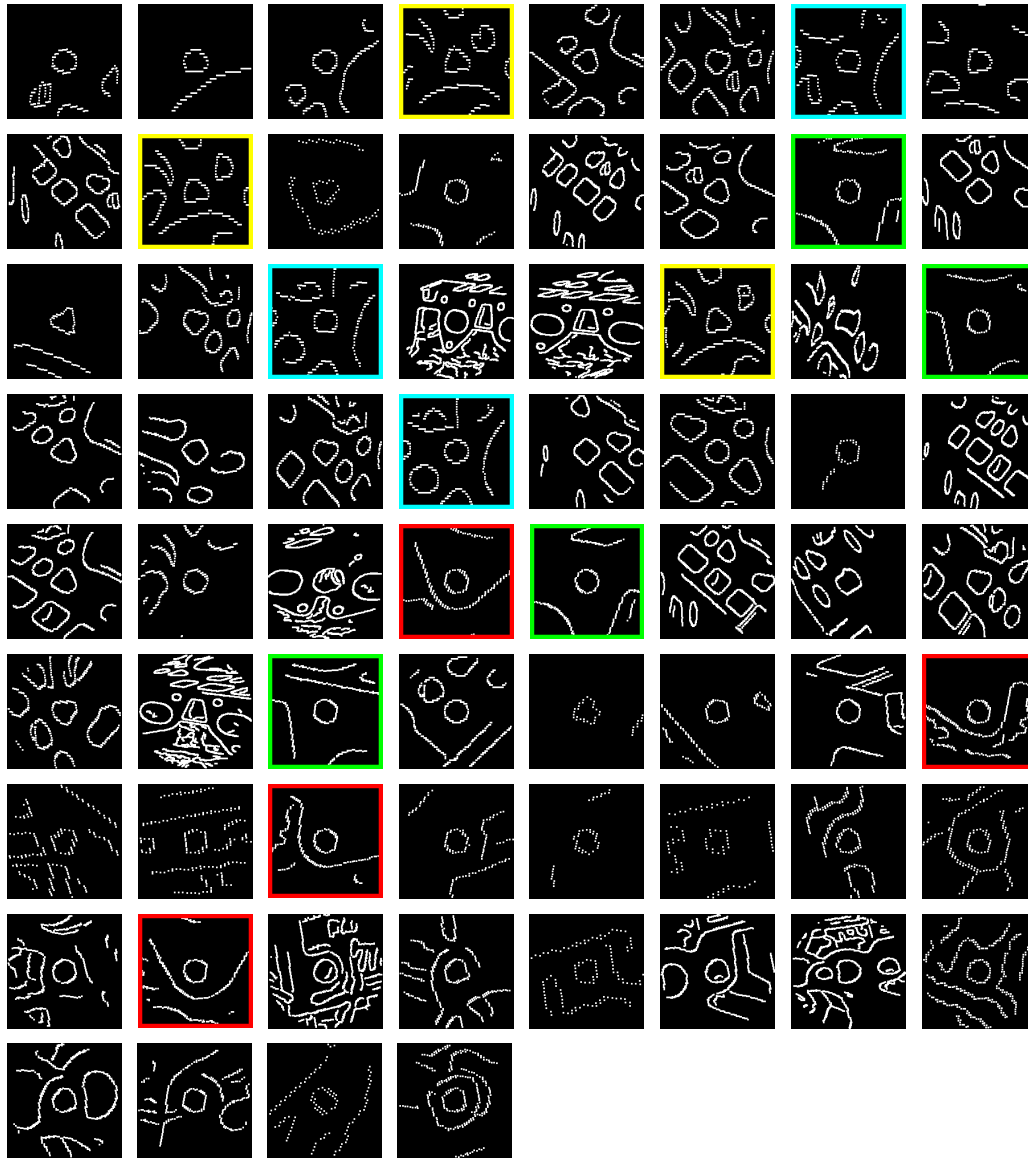


FIG. 3.7 – Les régions supports normalisées extraites de la première image d'apprentissage de la culasse. *Les régions supports encadrées d'une même couleur sont des exemples d'instances d'un même motif répétitif (même code couleur que l'image 3.6).*

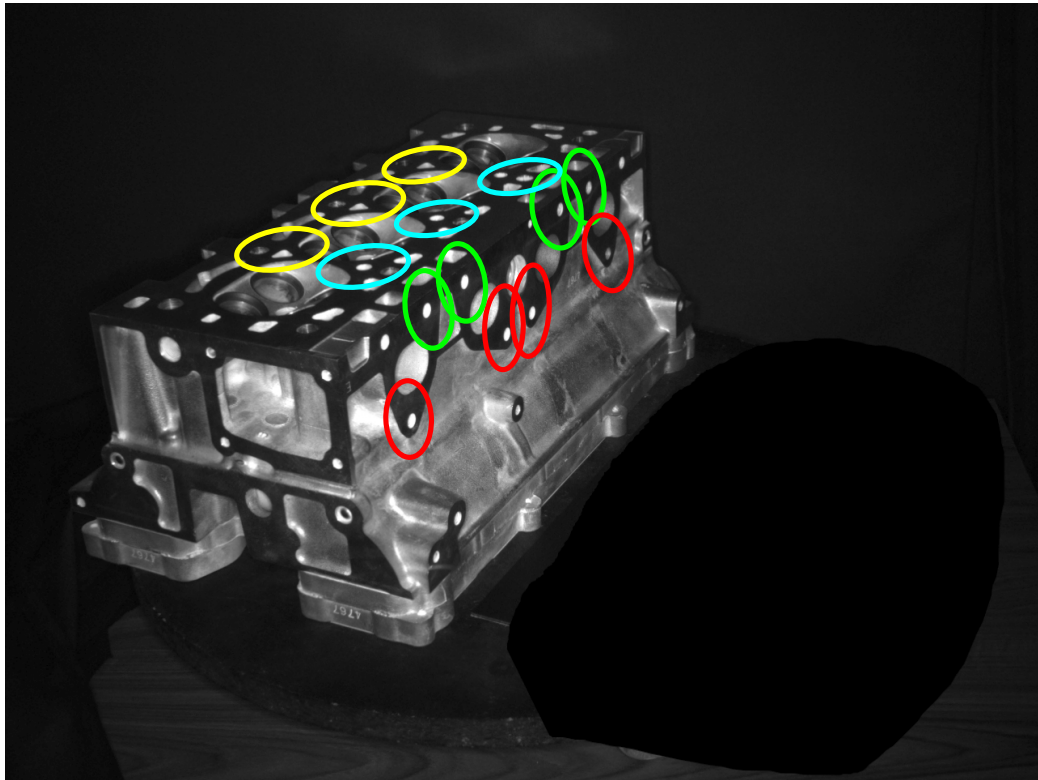


FIG. 3.8 – Seconde image d'apprentissage de la culasse. *Les ellipses de même couleur correspondent aux différentes régions supports d'un même motif répétitif. Les régions supports normalisées correspondantes peuvent être observées dans la figure 3.9.*

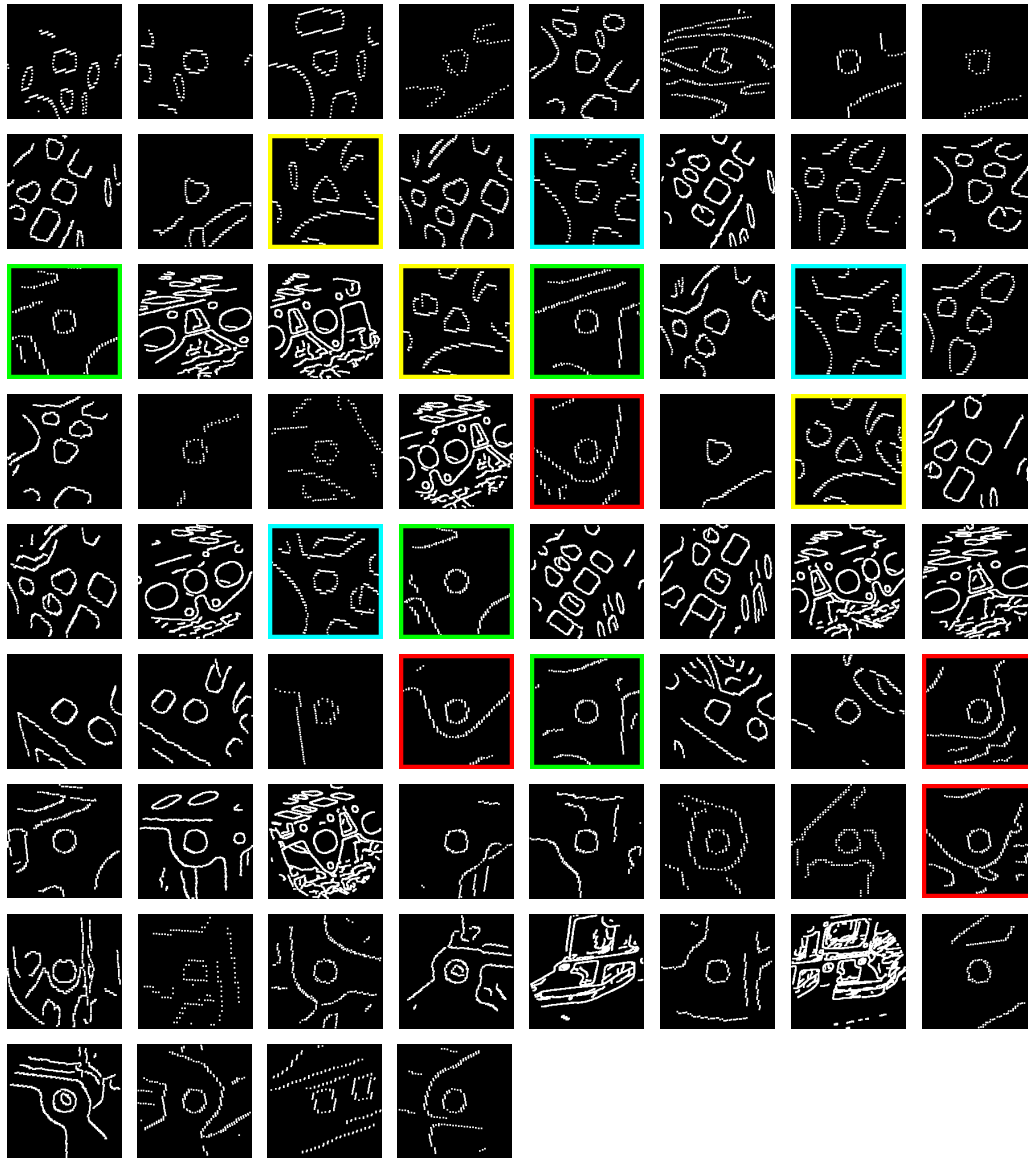


FIG. 3.9 – Les régions supports normalisées extraites de la seconde image d'apprentissage de la culasse. *Les régions supports encadrées d'une même couleur sont des exemples d'instances d'un même motif répétitif même code couleur que l'image 3.8).*

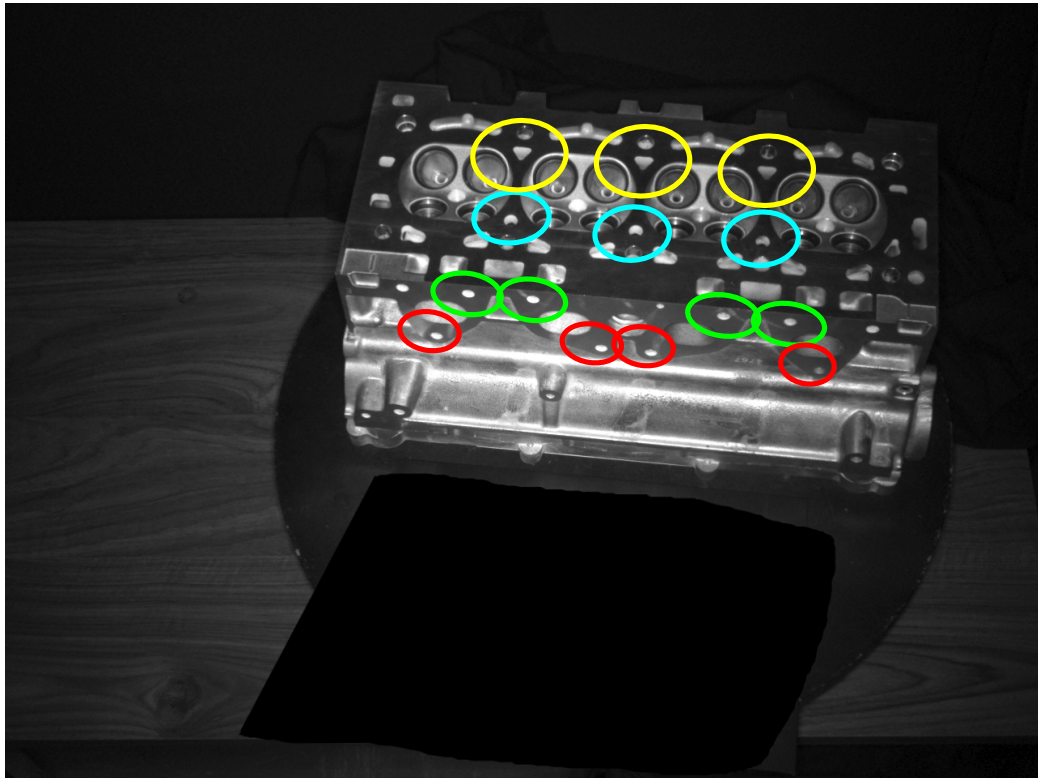


FIG. 3.10 – Troisième image d'apprentissage de la culasse. *Les ellipses de même couleur correspondent aux différentes régions supports d'un même motif répétitif. Les régions supports normalisées correspondantes peuvent être observées dans la figure 3.11.*

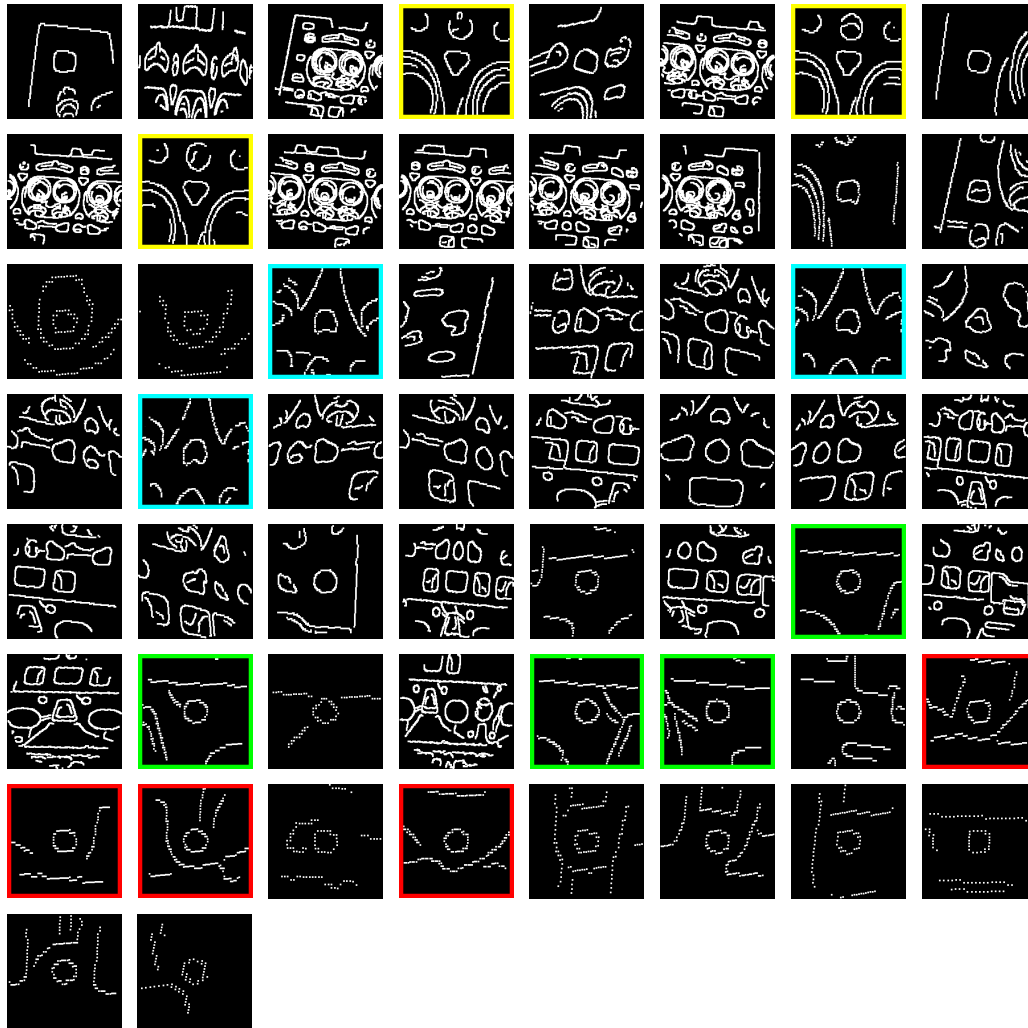


FIG. 3.11 – Les régions supports normalisées extraites de la troisième image d'apprentissage de la culasse. *Les régions supports encadrées d'une même couleur sont des exemples d'instances d'un même motif répétitif (même code couleur que l'image 3.10).*

Pour remédier à ce problème, la connaissance a priori de ces motifs répétitifs est intégrée au modèle. On introduit alors la notion de *descripteurs équivalents*.

Un descripteur M_k est équivalent à un descripteur M_j quand $D_{SC}(M_j, M_k) < \zeta_j$. L'ensemble des descripteurs du modèle équivalents à un descripteur M_j est noté $\{M_j^k\}$. Cet ensemble peut être déterminé de manière automatique lors de la création du modèle de l'objet. Il suffit alors d'étudier la distribution des distances entre M_j et les autres descripteurs du modèle pour établir cet ensemble. Les descripteurs équivalents correspondant alors aux descripteurs du modèle compatibles avec M_j (cf. figure 3.5). Cette étape étant réalisée hors ligne, elle ne pénalise pas les performances du processus de mise en correspondance.

Enfin, cette connaissance a priori peut alors être exploitée lors du processus de mise en correspondance. Chaque appariement $\langle I_i, M_{j(i)} \rangle$ est alors remplacé par un appariement multiple $\langle I_i, \{M_{j(i)}^k\} \rangle$ entre le descripteur image et l'ensemble des descripteurs équivalents à $M_{j(i)}$. Ainsi, le problème des erreurs d'appariements liées aux motifs répétitifs est remplacé par une ambiguïté sur l'identité des descripteurs images. Cette ambiguïté sera ensuite levée à l'aide des contraintes géométriques du modèle.

On note $\Psi = \{ \langle I_i, \{M_{j(i)}^k\} \rangle \}$ l'ensemble d'appariements obtenu à l'issue de l'étape de mise en correspondance basée sur l'apparence.

3.4 Évaluation

Le processus de mise en correspondance a été évalué sur les images d'une culasse de moteur automobile. Il s'agit d'un objet qui, d'une part, entre parfaitement dans le contexte de cette étude, et d'autre part, comporte de nombreux motifs répétitifs. Il s'agit donc d'un objet permettant d'illustrer les difficultés induites par les motifs répétitifs lors d'une mise en correspondance.

3.4.1 Protocole expérimental

Pour évaluer les performances du processus de mise en correspondance, une sphère de vues a été réalisée. Celle-ci contient 81 images dont la répartition spatiale est illustrée dans la figure 6.1 (page 115). Parmi ces 81 images, trois images ont été utilisées pour élaborer le modèle de l'objet selon la méthode décrite dans le chapitre 2. Les expériences consistent alors à établir, pour chacune des 81 vues, une mise en correspondance entre les descripteurs issus de l'image et ceux du mo-

dèle.

Les performances de la mise en correspondance sont évaluées en terme de précision et de rappel (cf. annexe A), le premier facteur représentant la proportion d'appariements corrects parmi les appariements sélectionnés à l'issue de la mise en correspondance, alors que le second représente le rapport entre le nombre d'appariements corrects sélectionnés et le nombre total d'appariements corrects pouvant être établis.

Pour chaque vue, les appariements ont été ordonnés selon leur qualité, c'est à dire en fonction de la distance entre la signature de *Shape Context* du descripteur image et celle du descripteur modèle auquel il est apparié. La vérité terrain étant connue pour chacune des 81 vues, la précision et le rappel ont pu être évalués. Les résultats présentés correspondent aux précisions et rappels pour des ensembles allant du seul meilleur appariement aux 100 meilleurs. La valeur de la précision et du rappel correspondent aux valeurs moyennes calculées sur l'ensemble des 81 vues.

D'autre part, afin d'évaluer le rôle des motifs répétitifs et de l'apport de la notion de descripteurs équivalents, trois expériences ont été réalisées.

La première expérience correspond au cas de figure où le problème des motifs répétitifs est ignoré. La mise en correspondance est alors réalisée sans exploiter la notion d'équivalence entre descripteurs. Les appariements issus de la mise en correspondance sont donc tous des appariements simples. Cette expérience sera désignée dans les courbes de résultats sous la légende "sans équivalence".

La seconde expérience correspond à la procédure de mise en correspondance proposée, c'est à dire avec une prise en compte de la notion d'équivalence. La mise en correspondance fournit donc des appariements multiples. Lors du calcul de la précision et du rappel, un appariement multiple comportant K hypothèses d'appariements est comptabilisé comme K appariements simples, chacun de ces appariements correspondant à l'une des K hypothèses. Cette expérience sera désignée dans les courbes de résultats sous la légende "avec équivalences".

Enfin, au cours de la troisième expérience, on suppose disposer d'un processus permettant de lever les ambiguïtés liées aux appariements multiples. Cette dernière expérience correspond donc au cas de figure où la notion d'équivalence a été prise en compte et où les ambiguïtés liées aux appariements multiples sont levées grâce à ce processus. Il s'agit d'illustrer l'apport que peut amener la notion d'équivalence lorsque celle-ci est correctement gérée. Lors du calcul de la précision et du rappel, un appariement multiple est considéré comme un appariement correct si l'une de ses hypothèses d'appariement est correcte, il est considéré comme un faux appariement dans le cas contraire. Cette expérience sera désignée

dans les courbes de résultats sous la légende "avec équivalences résolues".

3.4.2 Résultats

Le rappel

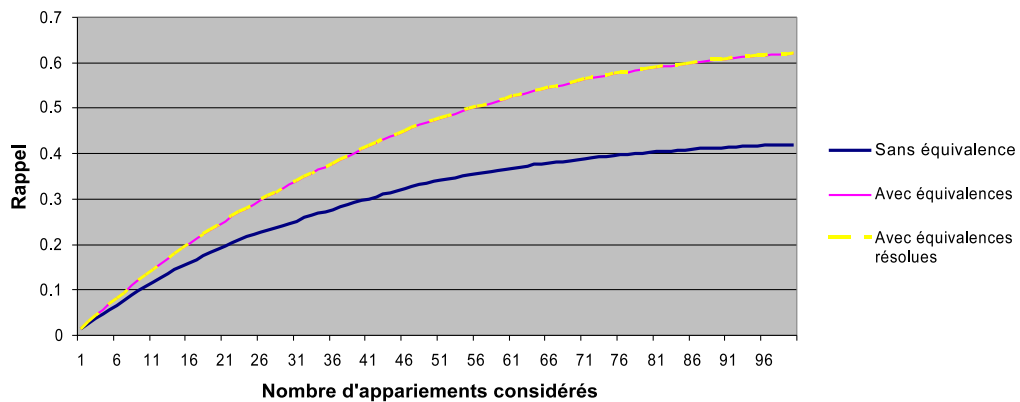


FIG. 3.12 – Rappel du processus de mise en correspondance selon l'apparence. *Le Rappel du processus de mise en correspondance est représenté en fonction du nombre d'appariements considérés (ces appariements étant classés selon la qualité de l'appariement, c'est à dire pour une D_{SC} croissant).*

Dans un premier temps considérons la courbe de rappel illustrée dans la figure 3.12. Tout d'abord, notons que les courbes de rappel des deux expériences considérant la notion d'équivalence sont identiques. Ce résultat est logique puisque seul le nombre d'appariements corrects est intégré au calcul du rappel et que celui est nécessairement le même dans les deux expériences (l'ajout du critère de décision sur l'identité des appariements multiples ne supprime aucun appariement correct).

Ensuite, si on compare la courbe de rappel de la première expérience à celle de la troisième, on constate que la notion d'équivalence permet d'obtenir un accroissement du rappel allant de 25% à 50%, avec gain moyen de l'ordre de 40%. Entre les deux expériences, la seule différence correspond à la comptabilisation des "faux" appariements correspondant à des appariements entre instances différentes d'un même motif répétitif. Dans le cas de la première expérience, de tels appariements sont considérés comme erronés alors qu'ils sont considérés comme corrects lors de la troisième expérience. La différence de rappel représente donc la quantité de faux appariements induits par les motifs répétitifs. Le gain de rappel

représente donc l'apport de la notion d'équivalence dans la gestion du problème des motifs répétitifs.

Ces premiers résultats permettent de conclure que la notion d'équivalence permet d'accroître de manière significative le nombre d'appariements corrects établis.

La précision

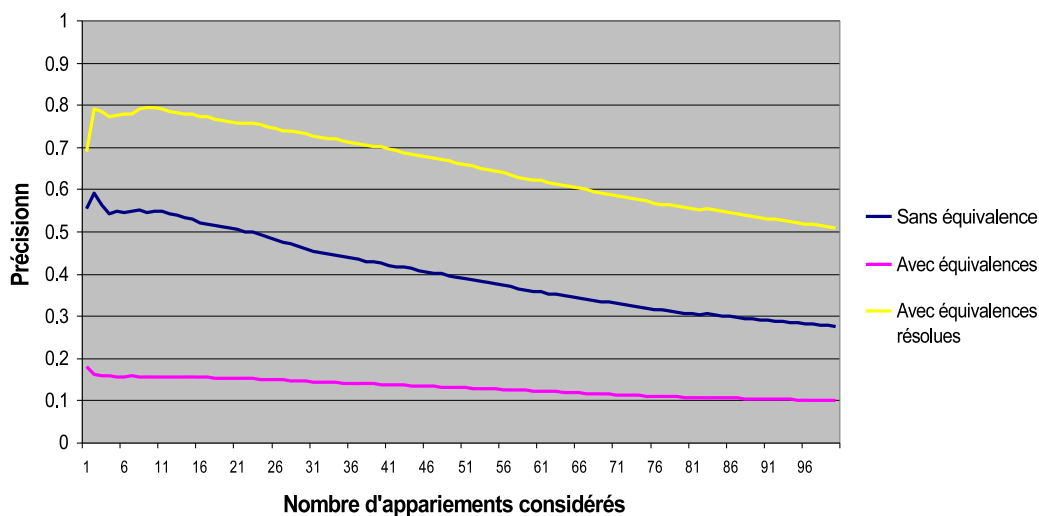


FIG. 3.13 – Précision du processus de mise en correspondance selon l'apparence. *La précision du processus de mise en correspondance est représentée en fonction du nombre d'appariements considérés (ces appariements étant classés selon la qualité de l'appariement).*

La courbe de précision est illustrée dans la figure 3.13. La précision correspond à un facteur de qualité primordial puisqu'elle représente la proportion d'appariements corrects que contient l'ensemble d'appariements fourni par le processus de mise en correspondance. Or, la proportion de faux appariement conditionne les performances d'un processus de calcul de la pose de l'objet. En effet, il est généralement considéré que la complexité d'un algorithme de calcul de pose robuste devient excessive lorsque celui-ci est appliqué à un ensemble d'appariements comportant plus de 50% de faux appariements. Par conséquent, pour qu'un calcul de pose robuste puisse être envisagé dans de bonnes conditions (temps de calcul non prohibitif), il est nécessaire que la précision du processus de mise en correspondance soit supérieure à 0.5.

Si on étudie la courbe de la première expérience, on constate que la précision du processus de mise en correspondance descend rapidement sous le seuil des 0.5. On peut donc considérer que le processus de mise en correspondance associé à l'expérience 1 n'offre pas une précision suffisante pour envisager un calcul de pose robuste en temps limité.

Si on étudie la courbe de la seconde expérience, on constate que la précision du processus de mise en correspondance ne dépasse pas 0.2, ce qui est nettement insuffisant pour envisager un calcul de pose. Si on compare les résultats de la première expérience à ceux de la seconde, on observe une chute de près de 70% de la précision. L'introduction de la notion d'équivalence entraîne donc une chute de la précision, ceci du fait de la présence de nombreux appariements multiples. En effet, lors de la seconde expérience, chaque hypothèse d'un appariement multiple a été considérée comme un appariement à part entière. Un appariement multiple contenant K hypothèses est alors équivalent à K appariements simples. Puisqu'il y a au mieux une hypothèse correcte parmi les K , la notion d'équivalence introduit nécessairement plus de faux appariements qu'elle n'introduit d'appariements corrects.

Si on considère les résultats de la dernière expérience, on constate que le processus de mise en correspondance offre une précision supérieure à 0.5, soit une précision permettant d'envisager un calcul de pose robuste dans des conditions satisfaisantes. D'autre part, on observe un gain de 30% à 80% de précision par rapport aux résultats de la première expérience, avec un gain moyen de l'ordre de 65%. Comme il était attendu, la notion d'équivalence, dès lors qu'elle est associée à un critère permettant de décider de la meilleure hypothèse pour chaque appariement multiple, augmente le nombre d'appariements corrects sans augmenter le nombre de faux appariements. Ainsi, si on compare les résultats de précision entre la seconde et la dernière expérience, on peut conclure qu'un processus de mise en correspondance exploitant la notion d'équivalence ne permet d'envisager un calcul de pose que si celui-ci est accompagné d'un processus permettant de gérer les appariements multiples.

La courbe de précision/rappel

La courbe de précision/rappel, telle qu'illustrée dans la figure 3.14, permet de lier les deux précédentes mesures. Pour plus de détails sur l'interprétation de cette courbe, on se référera à l'annexe A.

Cette courbe permet notamment de déterminer le rappel obtenu pour une précision de 0.5, soit la précision que l'on considère minimale pour envisager un calcul de pose. On constate alors que, pour cette précision, la première expérience

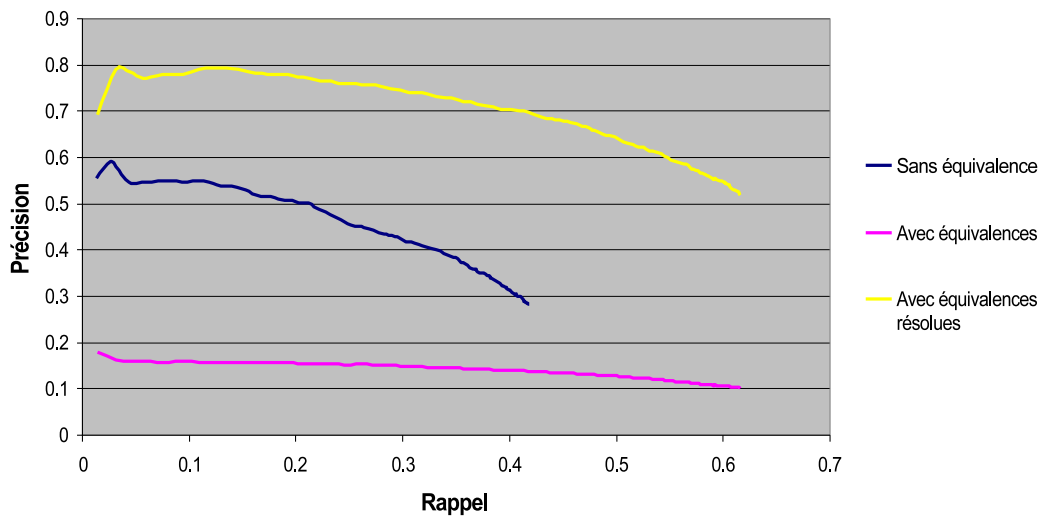


FIG. 3.14 – Courbe de précision/rappel obtenue sur les 100 premiers appariements du processus de mise en correspondance selon l'apparence.

atteint un rappel de 0.2 alors que la dernière atteint un rappel d'environ 0.6 (la seconde expérience ne permet pas d'atteindre une précision de 0.5). De ces résultats, on peut déduire que, sans notion d'équivalence et pour une précision permettant d'envisager un calcul de pose (i.e 0.5), le processus de mise en correspondance ne permet d'établir qu'une très faible fraction (20%) des appariements corrects possibles. Un système de recalage basé uniquement sur ce processus de mise en correspondance risque alors d'être à la fois peu robuste aux occultations et peu précis. A l'opposé, le processus de mise en correspondance de la troisième expérience permet d'établir près de 60% des appariements corrects possibles. Il permettrait donc un calcul de pose performant en terme de robustesse et de précision. Cependant, rappelons que ce processus nécessite une étape supplémentaire consistant à identifier la véritable identité à attribuer à chaque appariement multiple. En l'absence d'un tel processus, on obtient le processus de mise en correspondance de l'expérience 2 qui n'offre pas une précision suffisante pour envisager un calcul de pose.

3.5 Conclusion

Ce chapitre a permis d'introduire un processus de mise en correspondance selon l'apparence exploitant deux signatures qui caractérisant chacune des pri-

mitives du modèle. La complémentarité de ces signatures permet d'obtenir un processus à la fois rapide et discriminant.

Afin de pouvoir décider, à l'issue de cette mise en correspondance, si un appariement est correct ou non, un seuil sur la distance entre signature *Shape Context* a été proposé. Un processus d'apprentissage a été proposé afin d'établir, de manière automatique, un seuil propre à chaque descripteur du modèle. La valeur de ce seuil est établie de manière à permettre une certaine variabilité des signatures sans réduire significativement le pouvoir discriminant du descripteur.

Enfin, il a été montré que le précédent seuil permettait d'identifier automatiquement les motifs répétitifs présents au sein du modèle. La notion de descripteurs équivalents a alors été proposée afin de permettre une gestion de ces motifs lors de la mise en correspondance. Si cette notion d'équivalence permet d'accroître le nombre d'appariements corrects établis, celle-ci introduit des ambiguïtés au niveau des appariements. Comme l'a souligné une évaluation, sans une gestion efficace de ces ambiguïtés, la notion d'équivalence augmente de manière si significative le nombre de faux appariements qu'il est impossible d'envisager un calcul de pose. Cette évaluation a souligné la nécessité d'un processus permettant de lever ces ambiguïtés. Un tel processus est proposé dans le chapitre suivant.

Chapitre 4

Filtrage des appariements et calcul de la pose exploitant l'organisation géométrique des primitives

Objectif : L'étape finale d'un processus de recalage visuel consiste à estimer la pose de l'objet à partir des primitives visuelles identifiées lors de l'étape de mise en correspondance. Pour que cela soit possible, la proportion de faux appariements doit être suffisamment faible. L'objectif de ce chapitre est d'exploiter l'information géométrique 3D contenue dans le modèle de l'objet pour supprimer les faux appariements et calculer la pose de l'objet.

L'étape de mise en correspondance, présentée au chapitre précédent, exploite uniquement le modèle d'apparence sous-jacent au modèle mixte de l'objet. Le pouvoir discriminant des signatures caractérisant l'aspect d'une région support, permet de définir un processus à la fois rapide et performant. Cependant, la présence de nombreux motifs répétitifs dans l'objet induit de trop nombreuses ambiguïtés sur l'identité des primitives issues de l'image (appariements multiples) pour qu'une pose puisse être estimée efficacement.

Ce chapitre présente les deux étapes qui succèdent à la mise en correspondance selon l'apparence et aboutissent à l'estimation de la pose de l'objet. La première, présentée dans la section 4.1, correspond à un processus de filtrage des appariements qui vient achever le processus de mise en correspondance. Ce filtrage permet non seulement de lever les ambiguïtés d'identité sur les appariements, mais aussi de réduire le nombre de faux appariements. La seconde étape, expliquée dans la section 4.2, correspond au calcul de pose de l'objet à partir des appariements filtrés.

A la différence du chapitre précédent qui reposait exclusivement sur l'information d'apparence, ce chapitre exploite l'organisation géométrique 3D des primitives du modèle. Pour chacune des deux étapes présentées dans ce chapitre, deux approches sont proposées :

- la première correspond à une approche 2D, c'est à dire une approche où les contraintes de rigidité de l'objet sont exprimées dans l'espace image avant d'être utilisées ;
- la seconde correspond à une approche 3D, c'est à dire une approche où les contraintes de rigidité de l'objet sont directement exploitées dans l'espace objet (ou espace 3D).

4.1 Filtrage géométrique des appariements

Les solutions permettant de filtrer les faux appariements peuvent être regroupées en deux catégories : les *approches globales* et les *approches locales*.

Les *approches globales* [26, 4] visent à déterminer un unique ensemble d'appariements vérifiant la contrainte de rigidité de l'objet. Ces approches consistent généralement à déterminer la pose de l'objet pour ensuite identifier les faux appariements à l'aide de la contrainte de *viewpoint consistency*. Pour autant, le coût calculatoire de ces solutions est généralement considéré comme trop élevé lorsque la proportion de faux appariements dépasse les 50%.

Pour leur part, les *solutions locales* [83, 74, 44, 59] ne visent pas à déterminer un unique ensemble d'appariements cohérents. Ces méthodes étudient la cohérence de chaque appariement vis à vis de son voisinage. Le modèle ne caractérise plus uniquement un descripteur par sa signature mais aussi par l'identité des descripteurs appartenant à son voisinage et éventuellement leur organisation spatiale. Pour être cohérent avec le modèle, un descripteur image doit aussi posséder un voisinage semblable à celui du descripteur modèle auquel il est apparié. Puisque les approches locales ne travaillent que sur de très petits ensembles d'appariements, celles-ci peuvent supporter un taux de faux appariements dépassant les 50%. Néanmoins, les appariements étant sélectionnés un à un selon un critère très local, la cohérence globale de l'ensemble des appariements sélectionnés vis à vis du modèle n'est pas nécessairement assurée. D'autre part, ces contraintes de voisinage ne sont généralement pas assez discriminantes pour lever les ambiguïtés liées aux motifs répétitifs. En effet, comme on peut le voir dans la figure 3.1, ceux-ci partagent généralement un voisinage très similaire, voir même identique.

Dans cette section, nous proposons une solution intermédiaire à ces deux ap-

proches.

4.1.1 Principe de la méthode

La solution proposée se situe à la frontière des approches globales et locales. Tout comme l'approche globale RANSAC [26], la solution proposée s'appuie sur l'estimation de poses de l'objet. Pour que les performances ne soient pas pénalisées par la proportion de faux appariements, une pose est estimée à partir d'un seul appariement entre descripteur. En contrepartie, la pose estimée n'est que très approximative. Généralement, celle-ci n'approxime l'objet qu'au voisinage du descripteur utilisé lors du calcul de la pose. Par conséquent, une pose estimée à partir d'un appariement correct ne permet pas d'identifier la totalité des appariements corrects de Ψ , mais uniquement ceux situés dans ce voisinage. Il s'agit de l'aspect local de la méthode. Néanmoins, la taille du voisinage n'est pas prédéfinie contrairement aux approches locales classiques. En effet, la taille du voisinage est directement liée à la qualité de la pose. Ainsi, si la pose est de très bonne qualité, le voisinage peut couvrir la totalité de l'objet : on tend alors vers une approche globale. A l'opposé, si la pose n'est que très approximative, le voisinage correspondant est petit : on tend alors vers une approche très locale. Par conséquent, une seule pose approximative ne permet pas nécessairement d'identifier la totalité des appariements corrects. La solution consiste alors à utiliser non pas une seule pose comme dans une approche globale mais les N meilleures poses. Comme l'illustrent la figure 4.1, chacune de ces N poses permet d'approximer un voisinage différent de l'objet. En combinant l'ensemble de ces voisinages, on obtient une couverture quasi-totale de l'objet (cf. figure 4.1). Seuls les appariements cohérents avec l'une de ces N poses sont alors sélectionnés. Les différentes étapes de ce processus sont illustrées dans la figure 4.2

Dans les sections suivantes, ce processus de filtrage est détaillé. Tout d'abord, un processus permettant de calculer une pose approximative de l'objet à partir d'un seul appariement est présenté dans la section 4.1.2. Afin de déterminer l'ensemble des appariements cohérents avec chacune des poses estimées à partir d'un appariement, deux critères sont proposés dans la section 4.1.3. A partir de ces critères, un processus de tri des poses approximatives et de sélection des appariements à partir de celles-ci, est présenté dans la section 4.1.4.

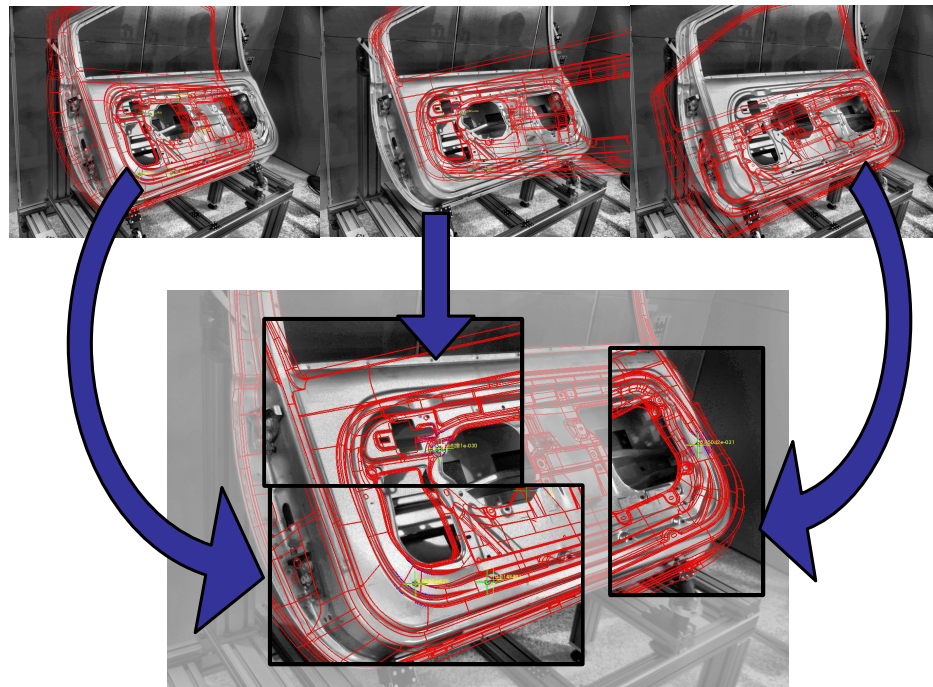


FIG. 4.1 – Principe général de la sélection des appariements selon un critère géométrique. *Les trois images du haut représentent 3 poses approximatives de l'objet, chacune estimée à partir d'un seul appariement. En combinant les voisinages sur lesquels ces poses approximent la pose réelle de l'objet (cadres noirs), on obtient une couverture quasi-complète de l'image de l'objet (image du bas).*

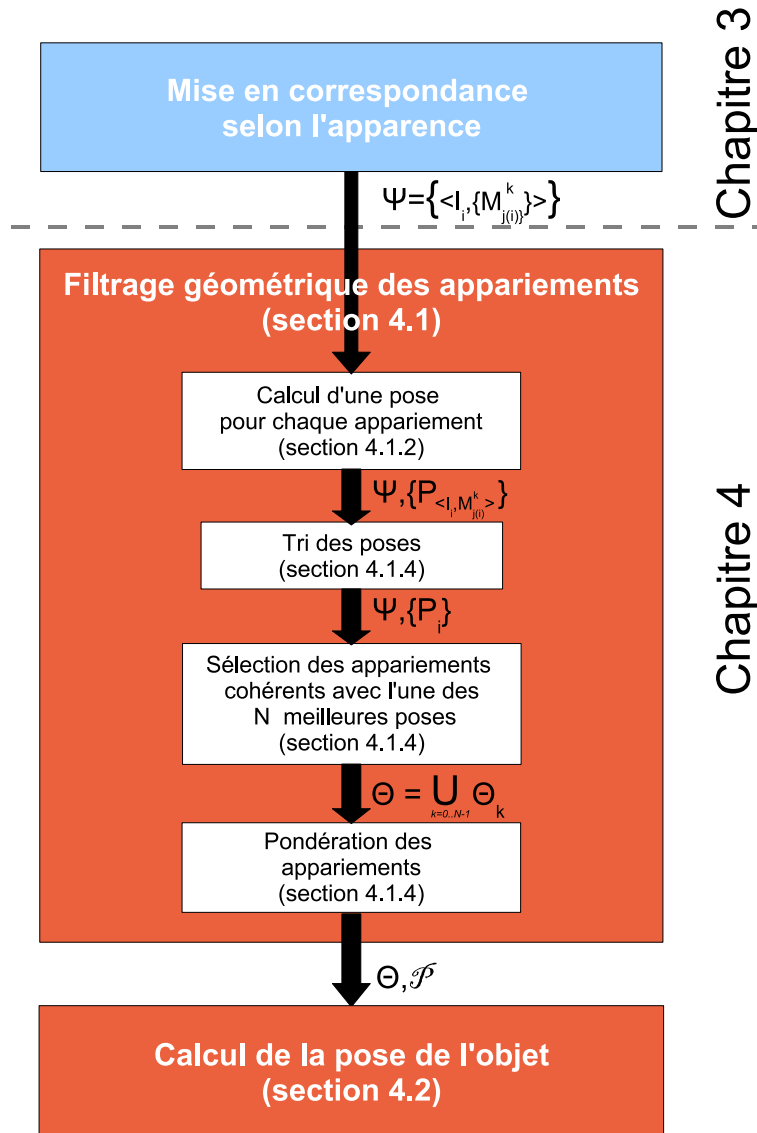


FIG. 4.2 – Les différentes étapes du filtrage géométrique.

4.1.2 Estimation d'une pose à partir d'un seul appariement

Grâce au motif 2D/3D, chaque descripteur du modèle est enrichi de suffisamment d'informations 3D pour qu'une pose puisse être estimée. Néanmoins, pour que le calcul de cette pose soit possible, il est nécessaire de déterminer la correspondance entre les points 3D du motif 2D/3D et ceux de l'image. La solution proposée consiste à établir une mise en correspondance approximative entre ces points, ceci afin d'obtenir une première évaluation de la pose. Cette pose est ensuite raffinée à l'aide d'un processus d'optimisation non linéaire.

Première estimation de la pose

Considérons un appariement $\langle I, M \rangle$ et notons :

- RS_I la région support associée au descripteur image I
- RSN_I la région support RS_I normalisée
- RSN_M la région support normalisée associée au descripteur du modèle M
- Soit \mathbf{F} la transformation affine, définie dans l'équation 2.3 (page 39), telle que $RSN_I = \mathbf{F}.RS_I$.

Si les régions supports correspondent à des surfaces planes de l'objet (hypothèse généralement admise dans le cadre des descripteurs locaux), alors il existe une homographie \mathbf{H} telle que :

$$RSN_M = \mathbf{H}.RS_I \quad (4.1)$$

La transformation \mathbf{H} permet donc d'apparier chaque point de contour de RSN_M à un point de l'image. Puisque les positions 3D des points de contour de la RSN_M sont données par le motif 2D/3D associé à M , il suffit de déterminer \mathbf{H} pour obtenir un nombre suffisant d'appariements 2D/3D pour l'estimation d'une pose.

Une solution approchée à ce problème consiste à supposer que \mathbf{H} peut être approchée par une affinité \mathbf{A} . On peut alors factoriser \mathbf{A} de la manière suivante :

$$\mathbf{A} = \mathbf{R}.\mathbf{F} \quad (4.2)$$

On peut alors reformuler l'équation 4.1 de la manière suivante :

$$RSN_M = \mathbf{H}.RS_I = \mathbf{R}.\mathbf{F}.RS_I = \mathbf{R}.RSN_I \quad (4.3)$$

L'ensemble de ces transformations est illustré dans la figure 4.3.

Puisqu'une RSN correspond à un disque unitaire, RSN_I et RSN_M sont deux disques unitaires reliés par une transformation affine \mathbf{R} . La transformation \mathbf{R} ne peut alors être qu'une rotation. La transformation \mathbf{F} ayant été estimée lors du calcul de la signature de I par l'équation (2.3), il suffit d'estimer la rotation \mathbf{R} pour

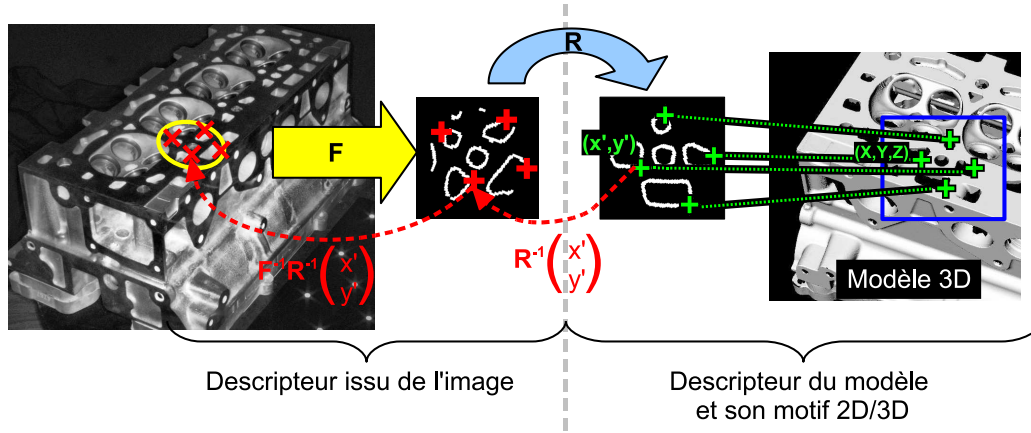


FIG. 4.3 – Lien entre les points 3D du motif 2D/3D et les points 2D de l'image.

obtenir \mathbf{A} .

Pour estimer \mathbf{R} , nous proposons d'exploiter une propriété du calcul de distance entre signatures *Shape Context*. En effet, l'estimation de \mathbf{R} peut être obtenue lors du processus de mise en correspondance avec le calcul de la distance $D_{SC}(s_I, s_M)$, où s_I et s_M sont les signatures respectives des descripteurs I et M . Cette distance consiste à chercher le maximum de corrélation pour des décalages circulaires successifs de s_I . Or, chaque permutation circulaire appliquée est équivalente à une rotation d'angle $2\pi/Q$ de la RSN_I . Si le maximum de corrélation est atteint après k permutations circulaires, alors le maximum de corrélation est obtenu pour une rotation d'angle $\theta = k.2\pi/Q$. On en déduit que \mathbf{R} correspond alors à une matrice de rotation d'angle θ .

\mathbf{F} et \mathbf{R} étant connues, la transformation affine $\mathbf{A} = \mathbf{R}.\mathbf{F}$ peut être calculée. La correspondance entre les points 3D du motif 2D/3D et les points 2D de l'image étant alors connue, une première estimation de la pose du descripteur peut être obtenue au moyen d'un algorithme de calcul de pose tel que POSIT [21].

Raffinement de la pose

Cependant, le fait que la transformation \mathbf{H} soit approximée par une affinité \mathbf{A} , mais aussi le fait que l'hypothèse de planarité de l'objet ne soit pas nécessairement vérifiée dans la pratique, introduisent des erreurs dans l'estimation de la position des points de l'image à appairer aux points 3D du motif 2D/3D. La pose

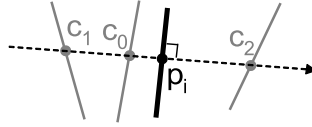


FIG. 4.4 – Optimisation de la pose. *En noir, la projection d'un contour. p_i correspond à la projection du point de contour P_i du modèle 3D. La flèche indique la direction de la normale au point de contour p_i . En gris, les contours de l'image. c_0 , c_1 et c_2 correspondent à 3 points de contour candidats pour le calcul d'erreur de reprojection.*

obtenue n'est donc pas parfaite, mais seulement approchée. Par conséquent, les points 3D du motif 2D/3D ne se projettent pas sur les points de contours de l'image sur lesquels ils sont censés se projeter. La distance entre la projection de ces points 3D et les contours de l'image peut être utilisée pour quantifier l'erreur de recalage. Nous proposons alors d'améliorer l'estimation de la pose au moyen d'un algorithme non linéaire consistant à minimiser cette erreur. On retrouve ce type de minimisation dans les méthodes de suivi d'objets telles que [32, 25, 82]

Notons \mathbf{P}_{init} la pose approximative estimée selon la méthode présentée au paragraphe précédent. L'optimisation de \mathbf{P}_{init} est obtenue en calculant un déplacement dans l'espace des poses permettant de superposer au mieux la projection des points 3D du motif 2D/3D avec les contours de l'objet dans l'image.

Pour estimer ce déplacement, un algorithme de minimisation non-linéaire est employé. Plus précisément, l'algorithme de Levenberg-Marquardt est utilisé pour minimiser l'erreur de reprojection du motif 2D/3D en ajustant les six paramètres de pose de la caméra. Cette erreur ε est calculée à partir des erreurs de reprojection de chacun des points 3D P_i du motif 2D/3D :

$$\varepsilon = \sum_i (\Delta(P_i))^2 \quad (4.4)$$

où $\Delta(P_i)$ correspond à l'erreur de reprojection du point P_i

L'estimation de $\Delta(P_i)$ nécessite de déterminer le point de contour c de l'image sur lequel P_i est censé se projeter. Notons p_i la projection de P_i selon une pose approximative. Cette pose étant proche de la solution, il est raisonnable de supposer que c se situe au voisinage de p_i . Cependant, à cause du "problème de l'ouverture" (*aperture problem*), la position de ce point ne peut être totalement déterminée. De ce fait, Harris [82] propose de limiter la zone de recherche à un segment centré en p_i et orienté selon la normale de p_i .

Cependant, plusieurs points de contours peuvent être détectés dans cette zone

de recherche. Il s'agit alors de déterminer celui qui sera utilisé au cours de la minimisation pour estimer l'erreur de reprojection.

Une solution proposée par Cippola et Drummond [25] consiste à utiliser le critère de proximité pour identifier le meilleur point de contour. Cependant, dans certaines situations (eg. contours proches les uns des autres, contours générés par une ombre,...), le point le plus proche n'est pas nécessairement le plus judicieux. De ce fait, Cippola et Drummond proposent d'associer aux appariements un poids proportionnel au nombre de points de contours détectés dans leur zone de recherche. Cette pondération est ensuite intégrée dans le processus de minimisation pour limiter l'influence des erreurs de reprojection dont les valeurs sont incertaines.

Une autre solution proposée par Vacchetti et al. [82] consiste à utiliser un estimateur robuste à hypothèses multiples. C'est cette solution qui est employée dans le cadre des travaux présentés. Si on considère un estimateur robuste ρ , l'estimateur à hypothèses multiples ρ^* qui lui est associé correspond à :

$$\rho^*(h_1, h_2, \dots, h_n) = \min_i \rho(h_i) \quad (4.5)$$

où les h_i correspondent aux différentes hypothèses. L'erreur de reprojection d'un point 3D du motif 2D/3D est alors définie de la manière suivante :

$$\Delta(P_i) = \rho^*(||c_1 - p_i||, ||c_2 - p_i||, \dots, ||c_n - p_i||) \quad (4.6)$$

où $||c_k - p_i||$ correspond à la distance euclidienne entre le point p_i et le $k^{\text{ème}}$ point de contour détecté dans la zone de recherche de p_i . Puisque le point p_i correspond au projeté de P_i , sa position dans l'image évolue au cours de l'optimisation. Par conséquent, le point de contour c_k le plus proche de p_i dépend des paramètres de pose de la caméra. L'identité du point de contour c_k le plus proche de p_i dépend donc de ces paramètres. Les estimateurs robustes à hypothèses multiples permettent de prendre en compte ces variations, ce qui n'est pas le cas des solutions qui fixent l'identité du point c_i correspondant à P_i avant l'optimisation.

On note $\mathbf{P}_{\langle I, M \rangle}$ la pose obtenue à partir de l'appariement simple $\langle I, M \rangle$ à l'issue du processus raffinement. Dans le cas d'un appariement multiple $\langle I_i, \{M_i^k\} \rangle$, une hypothèse de poses $\mathbf{P}_{\langle I_i, M_i^k \rangle}$ est générée pour chaque hypothèse d'identité de l'appariement. L'ensemble des poses générées par Ψ est noté $\Phi = \left\{ \mathbf{P}_{\langle I_i, M_i^k \rangle} \right\}$.

Évaluation

Cette évaluation vise à mesurer la précision d'une pose approximative estimée à partir d'un seul appariement.

Protocole expérimental L'expérience consiste à estimer un ensemble de poses approximatives à partir d'un ensemble d'appariements issus de données réelles¹. La vérité terrain est connue, l'erreur commise sur chacune des poses approximatives peut être mesurée. Cette évaluation a été réalisée sur trois objets : une culasse, une portière automobile et un objet étalon. Pour chacun de ces objets, les vues desquelles sont extraits les appariements, ainsi que le modèle de l'objet utilisé, sont présentés dans la section 6.1.1 (page 113).

L'erreur commise sur l'estimation de la pose de l'objet est caractérisée par trois erreurs :

- l'erreur d'orientation : il s'agit de l'angle² de la rotation permettant d'aligner le repère de l'objet réel et le repère de l'objet selon la pose estimée. Cette erreur est exprimée en degrés ;
- l'erreur de position globale : il s'agit de la distance euclidienne entre le centre de gravité de l'objet selon la pose réelle et le centre de gravité selon la pose estimée ;
- l'erreur de position locale : il s'agit de la distance euclidienne entre la position 3D du descripteur ayant servi au calcul de la pose et la position de ce même point selon la pose réelle.

La différence entre l'erreur de position locale et globale est illustrée dans la figure 4.5. Ces mesures permettent de comparer la précision globale de la pose à la précision locale de celle-ci. Le calcul de pose ayant été réalisé à partir de mises en correspondances localisées au niveau du descripteur, il n'est pas surprenant de constater que l'erreur de position soit plus faible au niveau du descripteur qu'au niveau du centre de gravité, comme l'illustre la figure 4.5.

Les résultats Les résultats obtenus pour chacune de ces erreurs et pour les trois objets sont résumés dans la figure 4.6. Les différentes erreurs sur les poses approximatives calculées sont classées par ordre croissant, ce qui permet de représenter la distribution des erreurs sur l'ensemble des poses. On peut alors voir que l'erreur de position locale médiane (ie. le centile 0.5) pour la portière est de 6

¹Cette expérience visant uniquement à caractériser la qualité du processus d'estimation d'une pose à partir d'un seul appariement, l'ensemble d'appariements considéré est dépourvu de faux appariements.

²On utilise une représentation axe/angle de la rotation.

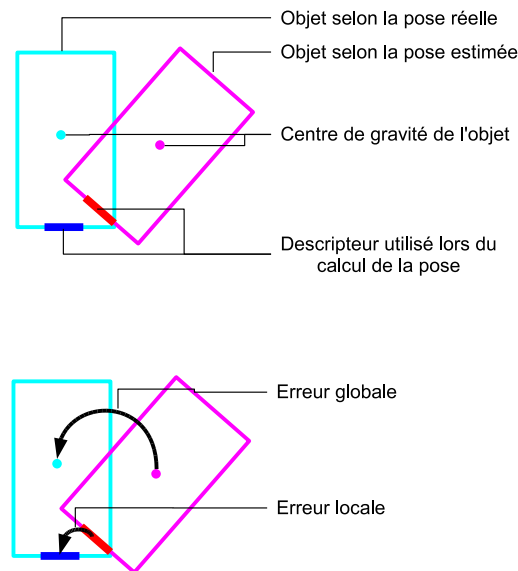
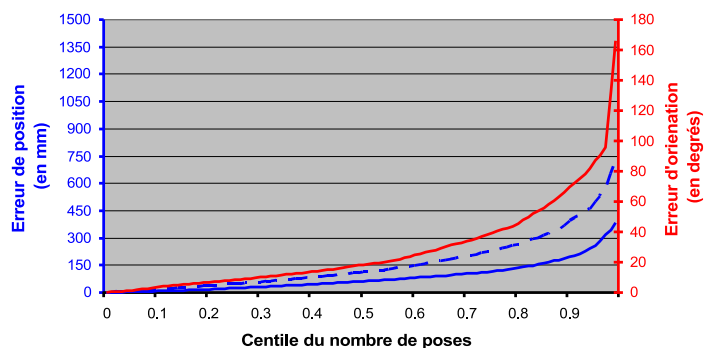


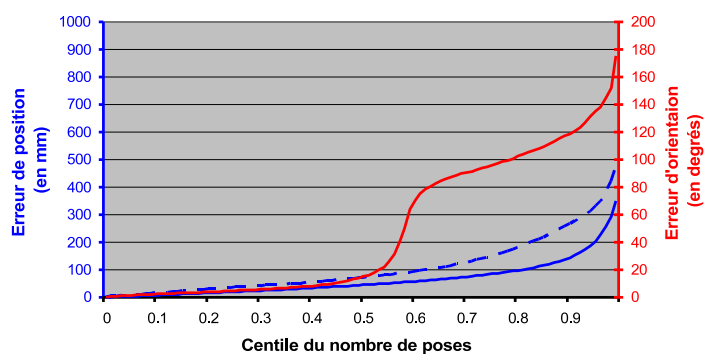
FIG. 4.5 – Les erreurs mesurées pour caractériser la qualité d'une pose estimée à partir d'un seul appariement

cm et que l'erreur d'orientation médiane de 17 degrés. D'autre part, l'échelle du graphique pour l'erreur de translation a été établie de manière à ce que l'erreur maximale représentable corresponde à la distance d'observation de l'objet. Il est ainsi possible de visualiser ce que représente l'erreur de position vis à vis de la distance d'observation.

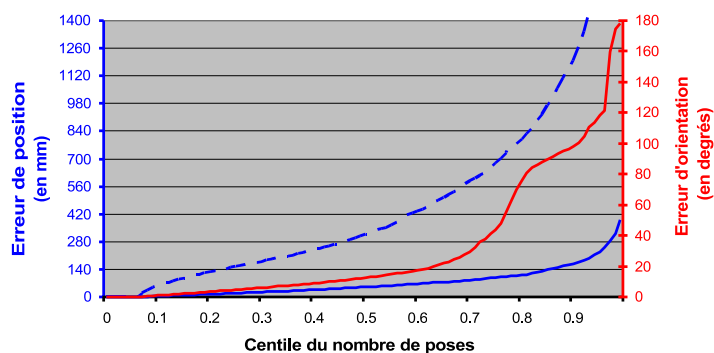
En ce qui concerne l'erreur d'orientation, on constate que près de la moitié des poses sont estimées avec une erreur comprise entre 10° et 20° , cela quelque soit l'objet considéré. S'agissant de l'erreur de position locale, on observe que celle-ci est inférieure à 10% de la distance d'observation pour plus de 80% des poses. On constate aussi que l'erreur de position globale est 1.75 fois plus élevée que l'erreur de position locale dans le cas de culasse, 2 fois plus élevée dans le cas de la portière et près de 6 fois plus élevée dans le cas de l'étalon. Ce dernier résultat confirme ce que la pose estimée est nettement meilleure au voisinage du descripteur employé pour la calculer. D'autre part, les variations du rapport entre l'erreur globale et l'erreur locale d'un objet à l'autre s'expliquent par l'effet de bras de levier. En effet, l'étalon est un objet très allongé alors que la culasse est un objet très compact. Or, plus l'objet est compact et plus le centre de gravité de l'objet est proche du descripteur employé pour le calcul de la pose et donc plus l'erreur globale se rapproche de l'erreur locale. A l'opposé, plus un objet est al-



(a) Résultats pour la portière



(b) Résultats pour la culasse



(c) Résultats pour l'étalon

FIG. 4.6 – Erreur sur la pose estimée à partir d'un seul appariement. *Erreur de position et d'orientation pour la portière, la culasse et l'étalon. La courbe bleue pleine correspond à l'erreur de position locale alors que la courbe bleue en pointillés correspond à l'erreur de position globale. La courbe rouge correspond à l'erreur d'orientation.*

longé et plus le centre de gravité de celui-ci est éloigné du descripteur utilisé pour le calcul de la pose, et donc plus l'erreur de position est amplifiée par l'effet levier.

Ces résultats montrent que la pose estimée à partir d'un appariement est avant tout une approximation de la pose de l'objet au voisinage du descripteur employé lors du calcul. Comme il sera montré dans les sections suivantes, l'identification des N meilleures de ces poses permet néanmoins de filtrer les faux appariements.

4.1.3 Critère de décision pour retenir les appariements cohérents avec une pose

Le processus de filtrage des appariements nécessite de déterminer un ensemble de poses approximatives de l'objet, et pour chacune d'entre elles identifier les appariements cohérents avec celle-ci. Le processus d'estimation de pose présenté dans les sections précédentes permet d'inférer cet ensemble de poses, une pose $\mathbf{P}_{\langle I, M \rangle}$ de l'objet étant estimée pour chaque appariement $\langle I, M \rangle \in \Psi$. Dans les paragraphes suivants, deux critères permettant de déterminer si un appariement $\langle \tilde{I}, \tilde{M} \rangle \in \Psi$ est cohérent avec une pose $\mathbf{P}_{\langle I, M \rangle}$ sont présentés. Le premier correspond à un critère 2D alors que le second correspond à un critère 3D.

Critère 2D

Pour déterminer si un appariement entre une primitive 3D d'un objet et une primitive 2D de l'image est cohérent vis à vis d'une pose \mathbf{P} , on dispose de la contrainte de *viewpoint consistency*. Cette contrainte correspond à une expression de la contrainte de rigidité de l'objet dans l'image. Elle stipule que les projections des primitives d'un objet doivent toutes être cohérentes avec un même point de vue. Par conséquent, si la pose de l'objet est connue, il est nécessaire que la primitive 3D du modèle se projette sur la primitive 2D de l'image à laquelle elle est appariée pour que l'appariement soit cohérent avec cette pose.

Dans la pratique, afin de compenser les erreurs d'estimation de la pose $\mathbf{P}_{\langle I, M \rangle}$, on considère que la primitive 3D du modèle doit seulement se projeter à proximité du descripteur image pour que l'appariement soit considéré cohérent avec la pose $\mathbf{P}_{\langle I, M \rangle}$. Notons $ErrProj_{\mathbf{P}_{\langle I, M \rangle}}$ l'erreur de reprojection d'un appariement pour une pose $\mathbf{P}_{\langle I, M \rangle}$:

$$ErrProj_{\mathbf{P}_{\langle I, M \rangle}}(\langle \tilde{I}, \tilde{M} \rangle) = \|(\mathbf{K}\mathbf{P}_{\langle I, M \rangle}\mathbf{G}_{\tilde{M}}) - \mathbf{g}_{\tilde{I}}\| \quad (4.7)$$

avec :

- \mathbf{K} : la matrice des paramètres intrinsèques de la caméra,
- $\mathbf{G}_{\tilde{M}}$: le point 3D à la position 3D du descripteur \tilde{M} ,
- $\mathbf{g}_{\tilde{I}}$: le centre de la région de \tilde{I} ,
- $|| \cdot ||$: la norme euclidienne.

C'est cette erreur de reprojection qui mesure le degré de cohérence d'un appariement avec une pose, et c'est un seuil sur cette erreur qui permet de décider si l'appariement est ou non cohérent avec cette pose. Un appariement $\langle \tilde{I}, \tilde{M} \rangle$ est alors considéré comme cohérent avec une pose P si :

$$ErrProj_{\mathbf{P}}(\langle \tilde{I}, \tilde{M} \rangle) < \delta \quad (4.8)$$

où δ correspond au seuil sur l'erreur de reprojection.

Le choix de la valeur de δ correspond à la principale difficulté de cette approche. En effet, sa valeur optimale dépend de la scène observée. Pour illustrer ce propos, considérons l'exemple d'un objet comportant un motif répétitif. Pour que l'erreur de reprojection permette de lever l'ambiguïté sur un appariement multiple, il faut que le seuil de l'erreur de reprojection soit fixé en fonction de la distance minimale à l'image entre deux instances du motif répétitif (on peut choisir le seuil comme étant égal à la moitié de cette distance par exemple). Or, cette distance minimale dépend de la distance d'observation de l'objet. Par conséquent, la valeur de ce seuil doit être décidée en fonction des conditions d'observations, ce qui implique un réglage manuel.

Une seconde limitation de cette approche est liée à la contrainte de *viewpoint consistency* : le respect de cette contrainte correspond à une condition nécessaire mais pas suffisante de la cohérence d'un appariement. En effet, deux primitives 3D différentes peuvent générer la même projection à l'image (cf. fig 4.8). En particulier, l'erreur de reprojection présentée ci-dessus ne prend en compte ni la taille du descripteur ni son orientation. On peut donc qualifier ce critère de *critère géométrique faible*. Pour obtenir un critère plus discriminant, une solution consiste à adopter une approche 3D.

Critère 3D

La principale limitation du critère 2D vient du fait que celui-ci travaille dans l'espace image. En effet, l'usage de la contrainte de *viewpoint consistency* implique une projection du modèle dans l'image pour pouvoir déterminer la cohérence de l'appariement. Cette projection introduit une perte de dimension (on passe d'un monde 3D à un monde 2D) et intègre la pose de la caméra dans le critère (ce qui rend le critère dépendant des conditions d'observations).

Afin d'éviter ces limitations, le critère 3D offre une solution équivalente au critère 2D mais travaillant dans l'espace objet et non dans l'espace image.

Considérons une hypothèse de pose $\mathbf{P}_{\langle I, M \rangle}$ estimée à partir de l'appariement $\langle I, M \rangle \in \Psi$ ainsi qu'un appariement $\langle \tilde{I}, \tilde{M} \rangle \in \Psi$ dont on veut savoir s'il est cohérent avec la pose $\mathbf{P}_{\langle I, M \rangle}$.

Puisque le modèle de l'objet définit la pose $[\tilde{\mathbf{R}}|\tilde{\mathbf{T}}]$ du descripteur \tilde{M} dans le repère de l'objet, il est possible de calculer la pose du descripteur \tilde{M} dans le repère de la caméra à partir de l'hypothèse de pose $\mathbf{P}_{\langle I, M \rangle}$ de l'objet. Notons $\mathbf{P}_{\langle I, M \rangle}^{\tilde{M}}$ la pose du descripteur \tilde{M} inférée à partir de la pose $\mathbf{P}_{\langle I, M \rangle}$. On a alors :

$$\mathbf{P}_{\langle I, M \rangle}^{\tilde{M}} = \mathbf{P}_{\langle I, M \rangle} \cdot [\tilde{\mathbf{R}}|\tilde{\mathbf{T}}] \quad (4.9)$$

Or, l'appariement $\langle \tilde{I}, \tilde{M} \rangle$ définit lui aussi la pose du descripteur \tilde{M} . En effet, le processus de calcul de pose présenté dans la section 4.1.2 permet d'estimer une hypothèse de pose $\mathbf{P}_{\langle \tilde{I}, \tilde{M} \rangle}$ de l'objet dans le repère de la caméra. Par conséquent, la pose $\mathbf{P}_{\langle \tilde{I}, \tilde{M} \rangle}^{\tilde{M}}$ du descripteur \tilde{M} peut être calculée à l'aide de l'équation 4.9.

Comme l'illustre la figure 4.7, on dispose alors de deux estimations de la pose de \tilde{M} : une première obtenue par l'hypothèse de pose $\mathbf{P}_{\langle I, M \rangle}$ de l'objet, la seconde obtenue à partir de l'appariement $\langle \tilde{I}, \tilde{M} \rangle$. Par conséquent, pour que cet appariement $\langle \tilde{I}, \tilde{M} \rangle$ soit cohérent avec la pose $\mathbf{P}_{\langle I, M \rangle}$, il est nécessaire que ces deux estimations de la pose de \tilde{M} soient identiques. Pour cela, notons Δ la transformation rigide reliant les deux estimations de la pose de \tilde{M} :

$$\Delta = \mathbf{P}_{\langle I, M \rangle}^{\tilde{M}} \cdot \mathbf{P}_{\langle \tilde{I}, \tilde{M} \rangle}^{\tilde{M}}^{-1}$$

Alors l'appariement $\langle \tilde{I}, \tilde{M} \rangle$ est cohérent avec la pose $\mathbf{P}_{\langle I, M \rangle}$ à condition que :

$$\Delta = Id_4$$

où Id_4 correspond à la matrice identité 4×4 .

Dans la pratique, les erreurs de calculs survenant dans l'estimation des poses $\mathbf{P}_{\langle I, M \rangle}$ et $\mathbf{P}_{\langle \tilde{I}, \tilde{M} \rangle}$ ne permettent pas que celles-ci soient parfaitement identiques. On considère alors que l'appariement $\langle \tilde{I}, \tilde{M} \rangle$ est cohérent avec la pose $\mathbf{P}_{\langle I, M \rangle}$ si la transformation rigide Δ est suffisamment faible. Si on note :

$$Err3D_{\mathbf{P}_{\langle I, M \rangle}}(\langle \tilde{I}, \tilde{M} \rangle) = d(\mathbf{P}_{\langle I, M \rangle}^{\tilde{M}}, \mathbf{P}_{\langle \tilde{I}, \tilde{M} \rangle}^{\tilde{M}})$$

où d correspond à une distance et où δ est un seuil sur cette distance, alors l'appariement $\langle \tilde{I}, \tilde{M} \rangle$ est cohérent avec $\mathbf{P}_{\langle I, M \rangle}$ si :

$$Err3D_{\mathbf{P}_{\langle I, M \rangle}}(\langle \tilde{I}, \tilde{M} \rangle) < \delta \quad (4.10)$$

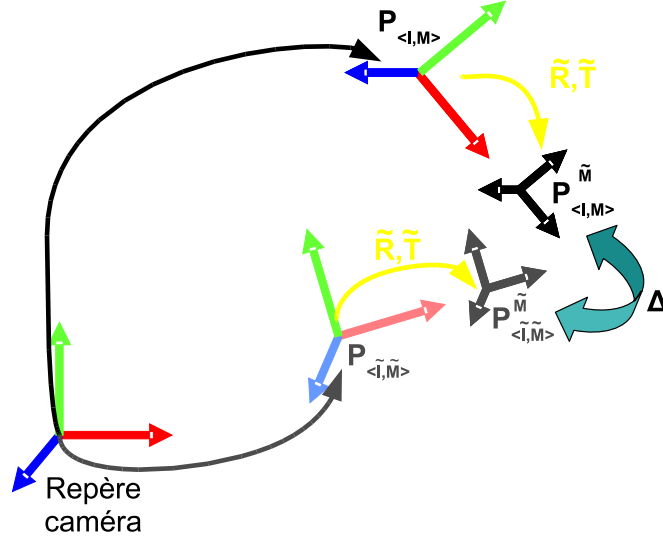


FIG. 4.7 – Critère de cohérence d'un appariement 3D. $\langle \tilde{I}, \tilde{M} \rangle$ est un appariement dont on veut savoir s'il est cohérent avec une hypothèse de pose $\mathbf{P}_{\langle I, M \rangle}$ de l'objet. $\mathbf{P}_{\langle \tilde{I}, \tilde{M} \rangle}$ correspond à la pose calculée à partir de l'appariement $\langle \tilde{I}, \tilde{M} \rangle$. La transformation rigide $[\tilde{R}|\tilde{T}]$, reliant le repère l'objet à celui de \tilde{M} , permet d'obtenir une estimation de la pose de \tilde{M} dans le repère de la caméra à partir de pour chacune des poses $\mathbf{P}_{\langle I, M \rangle}$ et $\mathbf{P}_{\langle \tilde{I}, \tilde{M} \rangle}$. Ces deux estimations, respectivement notées $\mathbf{P}_{\tilde{M}}^{\langle I, M \rangle}$ et $\mathbf{P}_{\tilde{M}}^{\langle \tilde{I}, \tilde{M} \rangle}$, sont reliées par une transformation rigide Δ . Pour que l'appariement $\langle \tilde{I}, \tilde{M} \rangle$ soit cohérent, Δ doit être suffisamment proche de Id_4 .

Afin d'obtenir un critère discriminant, la distance d doit intégrer dans une même mesure la composante translationnelle et rotationnelle de Δ . La distance retenue correspond à la distance des trièdres, notée *ErrTriedre* et présentée en annexe B. La prise en compte de ces deux composantes permet à ce critère d'être plus discriminant que le critère 2D comme l'illustre la figure 4.8.

On notera que tout comme pour l'approche 2D, la décision de la cohérence d'un appariement dépend d'un seuil. Cependant, ce seuil correspond à une distance dans l'espace objet. Par conséquent, ce seuil ne varie pas en fonction des conditions d'observations. Si on considère de nouveau l'exemple des motifs répétitifs, le seuil permettant de différencier deux instances d'un même motif répétitif doit être fixé en fonction de la distance 3D minimale entre deux instances de ce motif dans l'espace objet. Cette distance étant propre à la géométrie de l'objet, elle est intrinsèquement liée au modèle de l'objet et ne dépend donc pas des conditions d'observations.

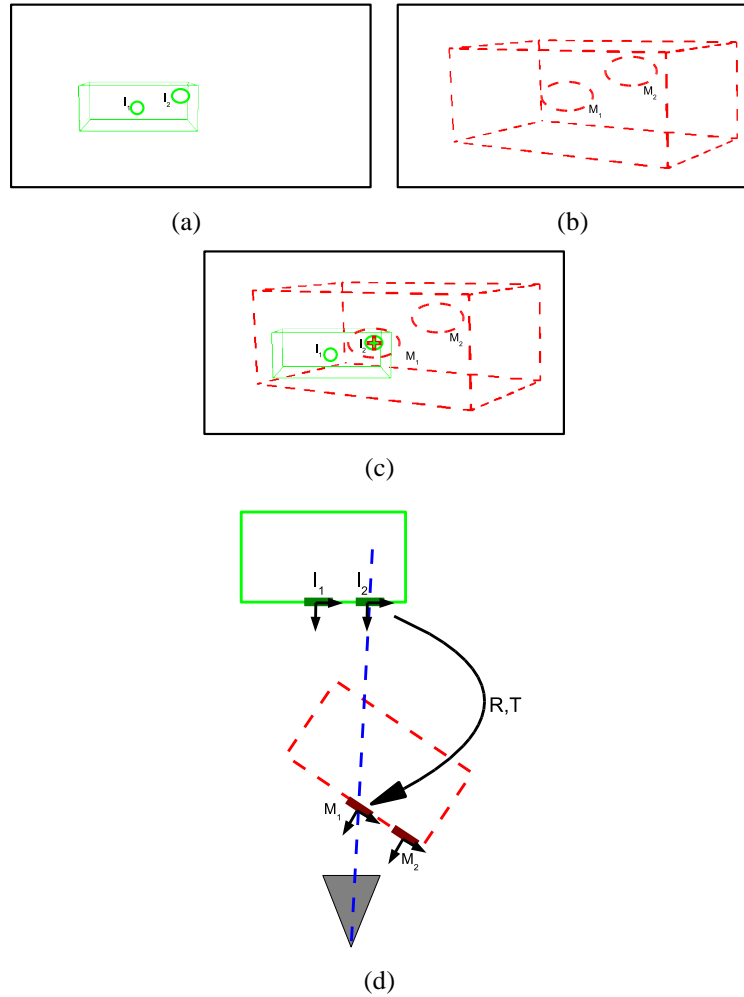


FIG. 4.8 – Comparaison des critères 2D et 3D pour l'identification d'appariements cohérents. (a) : l'image observée. Celle-ci représente un objet (parallélépipède vert) comportant deux descripteurs équivalents I_1 et I_2 (ellipses vertes). (b) : la scène selon une hypothèse de pose \mathbf{P} . La reprojection de l'objet selon \mathbf{P} est représentée en pointillés rouges. Les ellipses M_1 et M_2 sont les projections des descripteurs I_1 et I_2 . (c) : Superposition de (a) et (b). L'erreur de reprojection associée à l'appariement $\langle I_2, M_1 \rangle$ est nulle (la croix rouge se superpose à la croix verte). Bien qu'erroné, l'appariement est considéré comme cohérent pour le critère 2D. (d) : la même scène observée en vue de dessus. La transformation rigide $[\mathbf{R}|\mathbf{T}]$ reliant la pose du descripteur I_2 à celle du descripteur M_1 ne correspond pas à l'identité. Le critère 3D estime donc que cet appariement est incohérent avec la pose \mathbf{P} .

Par conséquent, l'approche 3D permet de fixer des seuils valides quels que soient les conditions d'observations. Ces seuils peuvent être établis hors ligne de manière automatique. Il est alors possible d'envisager un calcul de seuil spécifique pour chacune des primitives du modèle de l'objet. Ainsi, un seuil très bas est affecté pour les motifs répétitifs dont les instances sont proches alors qu'un descripteur qui n'est pas un motif répétitif se voit affecter un seuil bien plus élevé.

Dans la section suivante, nous présentons comment les poses approximatives permettent de filtrer les faux appariements à l'aide des critères de cohérences qui viennent d'être présentés.

4.1.4 Sélection des groupes d'appariements cohérents

Grâce aux critères présentés dans la section 4.1.3, il est possible d'identifier, pour chaque pose $\mathbf{P}_{\langle I_i, M_{j(i)} \rangle}$ calculée à partir d'un appariement $\langle I_i, M_{j(i)} \rangle \in \Psi$, l'ensemble des appariements de Ψ cohérents avec $\mathbf{P}_{\langle I_i, M_{j(i)} \rangle}$.

Cependant, de par sa construction, l'ensemble des poses $\Phi = \{\mathbf{P}_{\langle I_i, M_{j(i)} \rangle}\}$ calculées à partir de Ψ contient au moins autant de poses erronées que Ψ contient de faux appariements. Il est donc nécessaire de définir un processus permettant d'identifier ces poses.

La solution proposée consiste à ordonner les poses selon un critère de "qualité". A l'issue de ce tri, les appariements cohérents avec les meilleures poses sont sélectionnés. De plus, chaque appariement sélectionné se voit attribué une pondération proportionnelle à la probabilité que celui-ci soit un appariement correct.

Tri des hypothèses de poses

Le tri de l'ensemble des poses nécessite la définition d'un critère permettant de mesurer la qualité d'une pose. Le critère proposé évalue la cohérence de Ψ par rapport à la pose considérée. Pour que ce critère soit pertinent, celui-ci doit refléter le nombre d'appariements de Ψ cohérents vis à vis de la pose mais aussi la qualité de la cohérence des appariements. Ce critère doit aussi intégrer la notion d'appariements multiples et assurer que la qualité de la pose a été estimée sur un ensemble d'appariements cohérents (eg. deux appariements considérés comme cohérents ne peuvent pas partager le même descripteur modèle).

Dans un premier temps, considérons le cas simple où :

- Ψ ne comporte que des appariements simples : $\Psi = \{\langle I_i, M_{j(i)} \rangle\}$.

– les appariements de Ψ sont globalement cohérents, c'est à dire :

$$\forall \langle I_i, M_{j(i)} \rangle \in \Psi, \forall \langle I_n, M_{m(n)} \rangle \in \Psi, ((I_i = I_n) \vee (M_{j(i)} = M_{m(n)})) \Rightarrow i = n \quad (4.11)$$

Le critère se limite alors à estimer la somme des erreurs de cohérence des appariements. Afin d'être robuste aux valeurs aberrantes induites par les faux appariements, un estimateur robuste σ est appliqué à chaque erreur d'appariements. L'estimateur de Tukey est employé à cet effet. Cet estimateur nécessite cependant de fixer un paramètre au delà duquel toute erreur possède le même poids. La valeur de ce paramètre est alors fixée à δ , cette valeur correspondant à l'erreur au-delà de laquelle un appariement est considéré comme incohérent avec la pose. L'erreur $\varepsilon_P(\Psi)$ associée à Ψ pour la pose P correspond alors à :

$$\varepsilon_P(\Psi) = \sum_{\langle I_i, M_{j(i)} \rangle \in \Psi} \sigma(E(\langle I_i, M_{j(i)} \rangle), \delta) \quad (4.12)$$

où E est l'erreur de cohérence d'un appariement, celle-ci correspondant à l'erreur 2D *ErrRep* ou à l'erreur 3D *ErrTriedre* en fonction de l'approche retenue.

La pose qui minimise ε_P correspond alors à celle qui maximise le nombre d'appariements cohérents tout en minimisant leur erreur de cohérence, ce qui correspond bien au critère souhaité.

A présent, considérons un ensemble d'appariements Ψ quelconque. Dans ce cas, Ψ peut comporter des appariements multiples, du type $\langle I_i, \{M_{j(i)}^k\} \rangle$. La présence de ces derniers ne permet alors plus d'assurer la cohérence globale de l'ensemble des appariements de Ψ . Afin de pouvoir nous ramener au cas précédent, une seule hypothèse par appariement multiple doit être retenue, celle-ci étant sélectionnée de manière à assurer la cohérence globale de l'ensemble des appariements retenus. Notons $\Pi(\Psi)$ cet ensemble final d'appariements, où $\Pi(\cdot)$ correspond à la fonction de réalisant la sélection des appariements. L'erreur de cohérence de Ψ peut alors être définie de la manière suivante :

$$\varepsilon_P(\Psi) = \sum_{\langle I_i, M_{j(i)} \rangle \in \Pi_P(\Psi)} \varepsilon_P(\langle I_i, M_{j(i)} \rangle)$$

Cependant, la fonction Π ainsi définie n'est pas nécessairement unique. Néanmoins, les seules solutions pertinentes correspondent à celles pour lesquelles la cohérence de l'ensemble des appariements sélectionnés est maximale, c'est à dire celles pour lesquelles $\varepsilon_P(\Psi)$ est minimale. On ajoute donc à la fonction Π la contrainte de devoir minimiser $\varepsilon_P(\Psi)$.

En exprimant ce problème sous la forme d'un graphe biparti pondéré (le premier ensemble de sommets correspondant aux descripteurs images, le second correspondant aux descripteurs modèles, les arêtes étant définies par les appariements de Ψ et leur poids par ϵ_P), on s'aperçoit que la fonction Π n'est autre que la solution du *problème d'affectation optimale* (problème classique de la théorie des graphes). Parmi les différents algorithmes permettant de résoudre ce problème, nous avons retenu est l'algorithme Munkres [8], aussi connu sous le nom d'algorithme hongrois.

Sélection des appariements

Le critère de qualité d'une pose définie au paragraphe précédent permet d'ordonner les poses de Γ . Notons $\{\mathbf{P}_i\}_{i \geq 0}$ cet ensemble de poses ordonnées. Le filtrage des appariements consiste alors à extraire, pour chacune des N meilleures poses, le groupe d'appariements cohérents avec celle-ci.

Pour une pose \mathbf{P}_k , le groupe d'appariements Θ_k cohérent avec celle-ci est défini par :

$$\Theta_k = \{ \langle I_i, M_{j(i)} \rangle \text{ tel que } \langle I_i, M_{j(i)} \rangle \in \Pi_{\mathbf{P}_k}(\Psi) \text{ et } E_{\mathbf{P}_k}(\langle I_i, M_{j(i)} \rangle) < \delta \} \quad (4.13)$$

où E correspond à l'erreur de cohérence (ie. soit *ErrTriedre*, soit *ErrProj*) et δ correspond au seuil sur l'erreur de cohérence au delà duquel un appariement n'est pas considéré comme cohérent avec une pose.

L'ensemble d'appariements filtré correspond à l'ensemble d'appariements obtenus à partir de N groupes d'appariements issus des N meilleures poses, soit :

$$\Theta = \bigcup_{i=0..N} \Theta_k$$

La valeur N correspond à un paramètre du processus de filtrage. Dans les expériences du chapitre 6, la valeur de N est fixée à 15.

Pondération des appariements

Malgré la procédure de filtrage des appariements présenté au paragraphe précédent, il est possible que Θ comporte de faux appariements. Ces faux appariements sont généralement dus à la présence de poses erronées parmi les N meilleures poses ou au seuil trop lâche dans le critère de cohérence (δ trop élevé dans l'équation 4.13). Afin de limiter l'influence de ces faux appariements dans le reste du processus de recalage, une pondération est associée à chaque appariement de Θ . Cette pondération est établie de manière à refléter la probabilité qu'un

appariement soit correct.

Comme indiqué précédemment, une pose approximative $\mathbf{P}_{\langle I, M \rangle}$ issue d'un appariement $\langle I, M \rangle$ est généralement correcte pour une région de l'image centrée en I . Les descripteurs images étant répartis à la surface de l'objet, il est fréquent que les régions associées à deux poses $\mathbf{P}_{\langle I, M \rangle}$ et $\mathbf{P}_{\langle \tilde{I}, \tilde{M} \rangle}$ possèdent une intersection non nulle. Si ces deux poses sont correctes, la projection du modèle dans cette sous région selon $\mathbf{P}_{\langle I, M \rangle}$ ou $\mathbf{P}_{\langle \tilde{I}, \tilde{M} \rangle}$ est semblable. Par conséquent, si un appariement de cette sous région est considéré comme juste selon la pose $\mathbf{P}_{\langle I, M \rangle}$, il est fort probable qu'il le soit aussi selon la pose $\mathbf{P}_{\langle \tilde{I}, \tilde{M} \rangle}$. Si on note respectivement Θ_1 et Θ_2 les ensembles d'appariements cohérents avec $\mathbf{P}_{\langle I, M \rangle}$ et $\mathbf{P}_{\langle \tilde{I}, \tilde{M} \rangle}$, on a alors : $\Theta_1 \cap \Theta_2 \neq \emptyset$.

A l'opposé, si au moins l'une des deux poses $\mathbf{P}_{\langle I, M \rangle}$ ou $\mathbf{P}_{\langle \tilde{I}, \tilde{M} \rangle}$ est erronée, la projection du modèle dans cette zone de l'image diffère grandement selon la pose considérée. Par conséquent, il est peu probable que Θ_1 et Θ_2 partagent des appariements. On a donc généralement : $\Theta_i \cap \Theta_j = \emptyset$.

Par conséquent, il est d'autant plus probable qu'un appariement soit correct que celui-ci apparaît dans de nombreux Θ_i . Ainsi, la pondération w associée à un appariement $\langle I, M \rangle$ est définie par :

$$w(\langle I, M \rangle) = \text{Card}(\{\Theta_i \text{ tel que } \langle I, M \rangle \in \Theta_i\}) \quad (4.14)$$

Une densité de probabilité discrète \mathcal{P} , associée à Θ , peut alors être définie de la manière suivante :

$$\mathcal{P}(\langle I, M \rangle) = \frac{w(\langle I, M \rangle)}{\sum_{\langle \tilde{I}, \tilde{M} \rangle \in \Theta} w(\langle \tilde{I}, \tilde{M} \rangle)} \quad (4.15)$$

Le processus de filtrage proposé exploite la richesse en information géométrique du modèle pour supprimer les ambiguïtés liées aux motifs répétitifs et réduire le nombre de faux appariements. Une évaluation expérimentale de ce filtrage est proposée dans le chapitre 5. La section suivante présente un algorithme de calcul de pose exploitant les appariements filtrés.

4.2 Estimation Robuste de la pose

Le processus de filtrage fournit un ensemble d'appariements pondérés. Cet ensemble n'étant pas exempt de faux appariements, une pose ne peut être calculée à partir de celui-ci qu'à l'aide d'un algorithme robuste. De plus, pour atteindre des performances optimales, nous proposons d'intégrer la pondération des appariements dans le processus d'estimation de la pose.

L'algorithme proposé se décompose en deux étapes. La première consiste à estimer une pose initiale (section 4.2.1) qui est optimisée lors de la seconde étape (section 4.2.2).

4.2.1 Calcul de la pose initiale

Pour obtenir une première estimation \mathbf{P}_0 de la pose de l'objet à partir de Θ , un calcul de pose robuste basé sur l'algorithme RANSAC [27] est employé. Celui-ci a été préféré à une estimation de la pose selon la méthode des moindres carrés médians (ou LMEDS) en raison de la nature des pièces traitées. En effet, les pièces mécaniques présentent de nombreux motifs répétitifs qui sont généralement alignés. Un glissement du modèle, le long de l'un de ces axes de répétition, peut alors correspondre à une pose pour laquelle un grand nombre d'appariements sont cohérents, comme l'illustre la figure 4.9. Alors que tous les appariements sont erronés, un grand nombre d'entre eux peuvent présenter une erreur de reprojexion très faible. Il peut donc arriver que l'erreur médiane, sur ce type de pose, soit particulièrement faible alors que la pose est erronée. Il peut alors arriver que l'algorithme des LMEDS sélectionne une pose correspondant à un glissement du fait de sa très faible erreur médiane. Pour éviter ce problème, nous avons préféré utiliser un algorithme de calcul de pose robuste qui cherche à maximiser le nombre d'appariements corrects, à savoir l'algorithme RANSAC.

L'algorithme RANSAC se déroule en deux étapes répétées itérativement. Au cours de la première, un ensemble de N appariements³ est tiré au hasard parmi les appariements issus du processus de mise en correspondance. Ces N appariements sont alors utilisés pour estimer une pose. La seconde étape consiste à identifier l'ensemble des appariements cohérents avec la pose estimée : il s'agit de l'ensemble de consensus.

Ces deux étapes sont donc répétées jusqu'à ce qu'un ensemble de consensus de taille suffisante soit atteint ou que suffisamment d'itérations aient été réalisées. La

³Pour des raisons de performances, le valeur de N correspond généralement au nombre minimal d'appariements nécessaires au calcul d'une pose.

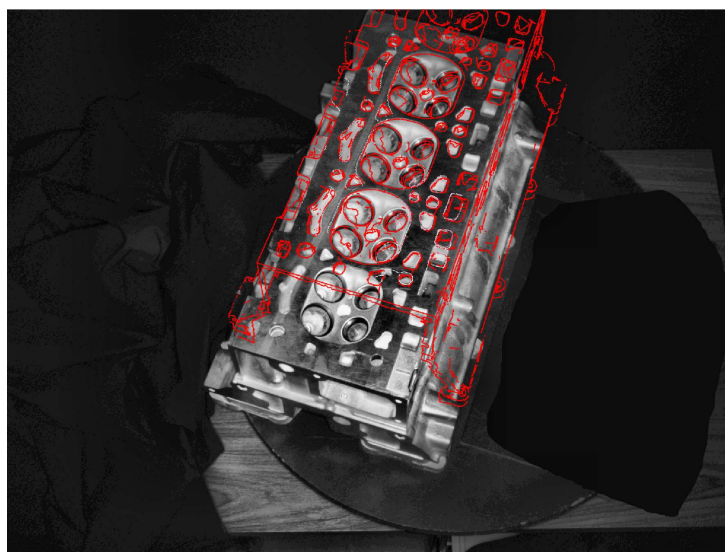


FIG. 4.9 – Phénomène de glissement du modèle CAO le long d'un axe de répétition des motifs répétitifs. *En rouge : la reprojection du modèle CAO de la culasse. On constate que le modèle CAO a glissé le long de l'axe longitudinal de la culasse. Bien que la pose soit erronée, le recalage n'est pas incohérent. On peut observer que les projections de nombreux alésages coïncident avec les contours de l'image.*

pose \mathbf{P}_0 de l'objet est alors calculée à partir de l'ensemble de consensus maximal obtenu.

Tirage aléatoire des appariements

Généralement, le tirage aléatoire est réalisé selon une densité de probabilité uniforme sur l'ensemble des appariements. Cependant, le processus de filtrage, présenté dans la section précédente, permet de déterminer, pour chaque appariement Θ , la probabilité que celui-ci soit un appariement correct. Réaliser le tirage aléatoire des appariements selon cette densité de probabilité \mathcal{P} permet d'augmenter la probabilité d'obtenir un échantillon constitué de N appariements corrects et par conséquent réduit le nombre d'itérations de l'algorithme.

Estimation du consensus

Le consensus est estimé à partir de la pose calculée à l'aide de l'algorithme POSIT [20] pour un échantillon d'appariements de taille $N = 4$.

Pour déterminer les appariements cohérents avec cette pose, l'ensemble de consensus est établi à l'aide de l'équation (4.13). La détermination de l'ensemble de consensus dépend donc du critère E_P choisi pour mesurer la cohérence d'un appariement avec une pose P , ce critère pouvant être l'un des deux critères proposés dans la section 4.1.3, c'est à dire :

- soit le critère 2D, $ErrReproj_P$, travaillant dans le repère image et exploitant la contrainte de *viewpoint consistency* (cf. équation 4.7) ;
- soit un critère 3D, $Err3D_P$, travaillant dans le repère objet et exploitant les contraintes géométriques 3D reliant les primitives du modèle de l'objet (cf. equation (4.10)).

L'ensemble des appariements de Θ cohérents avec la pose \mathbf{P}_0 est noté Φ :

$$\Phi = \{ \langle I_i, M_i \rangle \in \Theta \text{ tel que } \langle I_i, M_i \rangle \in \Pi_P(\Theta) \text{ et } E_P(\langle I_i, M_i \rangle) < \delta \} \quad (4.16)$$

4.2.2 Raffinement de la pose

Pour accroître la précision de la pose \mathbf{P}_0 , une optimisation des 6 paramètres de celle-ci est réalisée à l'aide d'un algorithme d'optimisation non-linéaire.

La fonction de coût à minimiser correspond au critère de qualité ε_P d'une pose P présenté dans la section 4.1 et défini par l'équation 4.12. Cette équation est reprise ci-dessous :

$$\varepsilon_P(\Psi) = \sum_{\langle I_i, M_i \rangle \in \Psi} \sigma(E(\langle I_i, M_i \rangle), \delta) \quad (4.17)$$

où Ψ correspond à l'ensemble des appariements issus de l'étape de mise en correspondance selon l'apparence.

Ce critère repose sur deux paramètres :

- le choix de l'erreur de cohérence E d'un appariement vis à vis d'une pose (ie. $ErrReproj$ ou $Err3D$) ,
- la valeur du seuil sur l'erreur de cohérence δ .

La position 2D d'un descripteur image étant connue avec plus de précision que sa position 3D, le critère retenu correspond au critère 2D. S'agissant du seuil δ , celui-ci est estimé à partir de la distribution des erreurs de reprojection des appariements de Θ cohérents avec \mathbf{P}_0 , c'est à dire Φ . Si on considère que cette distribution correspond à une distribution gaussienne centrée en 0 alors δ est fixé à trois fois l'écart type de cette distribution. Cette distribution n'étant cependant pas parfaitement gaussienne, l'écart type est remplacé par la déviation absolue médiane (ou MAD). Le seuil δ correspond alors à :

$$\delta = 3 \times 1.4826 \times \text{Mediane}(\{ErrReproj(\langle I_i, M_i \rangle) \text{ tel que } \langle I_i, M_{j(i)} \rangle \in \Phi\})$$

où le facteur 1.4826 permet au MAD d'obtenir une estimation sans biais de l'écart type lorsque la distribution respecte la loi normale.

La probabilité que l'erreur de reprojection d'un faux appariement soit inférieure au seuil δ est généralement négligeable. La fonction de coût peut donc intégrer un grand nombre de faux appariements sans que ceux-ci perturbent le processus d'optimisation. Par conséquent, la fonction de coût est évaluée sur Ψ , et non pas uniquement Θ , ceci afin d'intégrer l'ensemble des appariements corrects au processus d'optimisation.

4.3 Conclusion

Ce chapitre a permis de montrer comment l'information géométrique, contenue dans le modèle de l'objet, peut être exploitée pour filtrer les faux appariements et calculer une pose précise de l'objet.

Tout d'abord, un processus de filtrage géométrique des appariements a été proposé. Celui-ci permet non seulement de filtrer les faux appariements mais aussi de lever les ambiguïtés liées à la présence de motifs répétitifs. Ce processus repose sur le lien établi par le modèle entre l'apparence de l'objet et sa géométrie. Ce lien repose sur le motif 2D/3D et permet d'estimer une pose approximative de l'objet à partir d'un seul appariement. Il est alors possible d'estimer rapidement une pose pour chaque appariement issu du processus de mise en correspondance selon l'apparence. Ces poses sont ensuite exploitées pour identifier les groupes d'appariements cohérents avec les contraintes de rigidité du modèle 3D de l'objet. Deux solutions distinctes ont été proposées pour cette étape : la première exprime les contraintes de rigidité dans l'espace image avant de les utiliser (le critère 3D), alors que la seconde exploite ces contraintes directement dans l'espace objet (le critère 3D). Enfin, une pondération des appariements a été proposée afin de réduire l'influence des faux appariements qui n'ont pas été supprimés.

Dans un dernier temps, un algorithme de calcul de pose à la fois robuste et précis a été proposé. Cet algorithme exploite les appariements pondérés issus du filtrage géométrique pour obtenir une première estimation de la pose de l'objet. Une optimisation de cette pose est ensuite réalisée pour accroître la précision du recalage.

Les deux chapitres suivants sont deux chapitres d'évaluation expérimentale.

Le premier évalue les performances du processus de filtrage géométrique alors que le second évalue les performances du processus de recalage dans son ensemble.

Chapitre 5

Evaluation du filtrage géométrique

Ce chapitre présente une évaluation expérimentale du processus de filtrage géométrique des appariements. Il s'agit d'évaluer la capacité de celui-ci à supprimer les faux appariements et à lever les ambiguïtés d'identité des appariements multiples. Il s'agit aussi d'évaluer l'influence du critère de cohérence d'un appariement sur la qualité du filtrage.

Ce chapitre débute avec la description du protocole expérimental (section 5.1). Une première série d'expériences est ensuite présentée dans la section 5.2. Celle-ci vise à déterminer lequel des critères de cohérences 2D et 3D est le plus pertinent. La section 5.3 correspond à une seconde série d'expériences évaluant les performances du processus de filtrage.

5.1 Protocole expérimental

5.1.1 Les mesures

La qualité du filtrage est évaluée au moyen de deux mesures :

- la précision pondérée : il s'agit de la mesure de précision (cf. annexe A) dans laquelle la probabilité \mathcal{P} associée à chaque appariement de Θ est prise en compte.

$$\text{Précision}(\Theta) = \sum_{\langle I_i, M_{j(i)} \rangle \in T_\Theta} \mathcal{P}(\langle I_i, M_{j(i)} \rangle) \quad (5.1)$$

où $T_\Theta \subseteq \Theta$ correspond à l'ensemble des appariements corrects contenus dans Θ ;

- le rappel : il s'agit de la mesure de rappel classique (cf. annexe A) qui, dans le cadre de cette évaluation, correspond au nombre d'appariements

corrects retenus par le processus de filtrage par rapport au nombre total d'appariements corrects contenus dans Ψ , soit :

$$\text{Rappel}(\Theta) = \frac{\text{Card}(T_\Theta)}{\text{Card}(T_\Psi)} \quad (5.2)$$

où $T_\Psi \subseteq \Psi$ correspond à l'ensemble des appariements corrects contenus dans Ψ ;

5.1.2 Les données

Dans les différentes expériences, le processus de filtrage géométrique est appliqué à différents ensembles d'appariements Ψ issus de mises en correspondances selon l'apparence. Ces ensembles d'appariements Ψ sont tous issus de processus de mises en correspondances réel entre un modèle d'objet et les différentes images d'une sphère de vue de celui-ci.

Les évaluations portent sur trois objets mécaniques : une culasse automobile, une portière automobile ainsi qu'un objet étalon. Les caractéristiques de ces objets, leur modèle, ainsi que les sphères de vues qui les accompagnent, sont présentés dans le chapitre 6.

5.1.3 Les expériences

Les performances du processus sont évaluées pour deux séries d'expériences. La première évalue la pertinence du critère de cohérence. La seconde évalue l'influence du nombre de poses utilisées pour former Θ sur la précision et le rappel du processus.

Au cours de ces différentes expériences, les deux critères de cohérences présentés dans la section 4.1.3 sont évalués, c'est à dire le critère 2D et le critère 3D. Au cours des expériences, un appariement est rejeté dès lors que l'erreur sur le critère de cohérence dépasse :

- 70 pixels d'erreur de reprojection pour le critère 2D,
- $1/5^{\text{ème}}$ de la longueur de l'objet en erreur de translation, ou 70° d'erreur d'orientation, pour le critère 3D.

5.2 Pertinence du critère de cohérence

Au cours du processus de filtrage des appariements, le rôle de ce critère de cohérence est double : d'une part, il permet d'évaluer la qualité des poses, d'autre part il permet de sélectionner les appariements cohérents avec l'une des N meilleures poses. Afin d'évaluer l'influence du critère sur ces différentes étapes, la

qualité de chaque ensemble d'appariements Θ_i est estimée en terme de précision et de rappel. Les résultats présentés correspondent donc à la précision et au rappel du processus de filtrage réalisé à partir de chacune des poses approximatives \mathbf{P}_i , les résultats étant présenté selon l'ordre des \mathbf{P}_i .

L'expérience consiste donc à évaluer $Rappel(\Theta_i)$ et $Précision(\Theta_i)$ avec i allant de 0 à $Card(\Psi)$. Cependant, afin d'améliorer la lisibilité des courbes de résultats, on présente ici les résultats des 200 premières poses. Pour un objet donné, ces mesures sont effectuées pour chacune des images de l'objet considéré. Les résultats présentés correspondent aux valeurs moyennes et médianes de ces mesures sur l'ensemble des images traitées.

5.2.1 Le rappel

La figure 5.1 présente les résultats obtenus pour la portière. D'une manière générale, on constate tout d'abord que les différentes courbes sont décroissantes, ce qui indique que les poses \mathbf{P}_i sont bien ordonnées selon la quantité d'appariements corrects qui leur sont cohérents. On peut aussi noter que quelque soit le critère employé, le rappel est supérieur à 0.85 pour les premiers Θ_i , ce qui indique que plus de 85% des appariements corrects sont identifiés dans chacun des Θ_i . Ce résultat indique que lorsque les poses \mathbf{P}_i approximent de manière suffisamment fine la pose réelle de l'objet, les critères de cohérence permettent d'identifier la quasi-totalité des appariements corrects. Ceci reflète l'aspect global du processus de filtrage évoqué dans la section 4.1.1. Le fait que le rappel décroisse pour les Θ_i suivants pour tendre vers un rappel nul souligne le fait que ce processus de filtrage n'est pas toujours global, ce qui le situe à l'intermédiaire des approches de filtrages géométriques globales et locales.

D'autre part, on peut observer que le critère 2D fournit un meilleur rappel que le critère 3D. Le critère 2D permet donc d'identifier un plus grand nombre d'appariements corrects que le critère 3D. Cette différence s'explique principalement par le fait que le critère 3D n'est vérifié par un appariement correct $\langle I_i, M_i \rangle$ que si la pose $\mathbf{P}_{\langle I_i, M_i \rangle}$ calculée à partir de celui-ci est suffisamment précise. Par contre, pour que ce même appariement vérifie le critère 2D, il suffit que la position dans l'image du centre du descripteur I_i soit correctement déterminée. Or, le calcul du centre d'une région support est un processus relativement simple et fiable, ce qui n'est pas le cas du calcul de la pose. Par conséquent, il est plus fréquent qu'un appariement correct ne vérifie pas la critère 3D que ce n'est le cas pour le critère 2D.

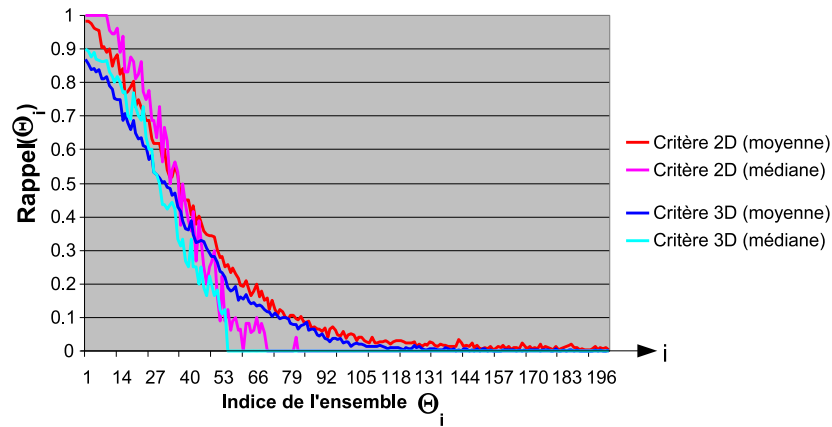


FIG. 5.1 – Première expérience : moyenne et médiane de $\text{Rappel}(\Theta_i)$ calculées pour les 50 images de la portière.

Les résultats qui viennent d'être présentés sont confirmés par ceux obtenus lors des expériences menées sur la culasse (figure 5.4) et l'étalon (figure 5.5).

5.2.2 La précision

La figure 5.2 présente les résultats de précision obtenus avec la portière.

On constate tout d'abord que les courbes de précision sont décroissantes, ce qui confirme le fait que les critères de cohérence prennent à la fois en compte le nombre d'appariements cohérents (ce qui correspond à un rappel élevé) mais aussi la qualité de cette cohérence (ce qui correspond à une précision élevée). On peut noter que cette précision varie entre 0.8 et 1 pour les premiers ensembles Θ_i , ce qui signifie que plus de 80% des appariements sélectionnés sont des appariements corrects. Sachant que le rappel atteint approximativement 0.9 pour ces mêmes Θ_i , cela signifie que ces ensembles d'appariements permettent d'identifier près de 90% des appariements corrects pour un taux de faux appariements de l'ordre de 20%. Les premières poses \mathbf{P}_i suffisent donc à filtrer efficacement les appariements.

On peut aussi observer que pour les ensembles Θ_i suivants, la précision décroît pour tendre vers 0. Cela signifie que la proportion de faux appariements augmente avec la dégradation de la qualité des poses. Cependant, cette baisse de la précision est conjuguée à une baisse du rappel. Le nombre d'appariements corrects sélectionnés devenant plus faible, le poids d'un faux appariement dans le calcul de la précision augmente. En particulier, lorsque le rappel est nul, la précision l'est nécessairement aussi. Dans ce cas, la précision et le rappel ne permettent alors plus de déterminer la quantité de faux appariements sélectionnés, celle-ci pouvant

être nulle comme très importante. Pour déterminer la robustesse d'un critère de cohérence aux poses erronées, il est intéressant de mesurer la quantité de faux appariements que celui-ci sélectionne à partir d'une pose erronée. Afin d'évaluer cet aspect, la mesure de co-rappel sera proposée dans la suite.

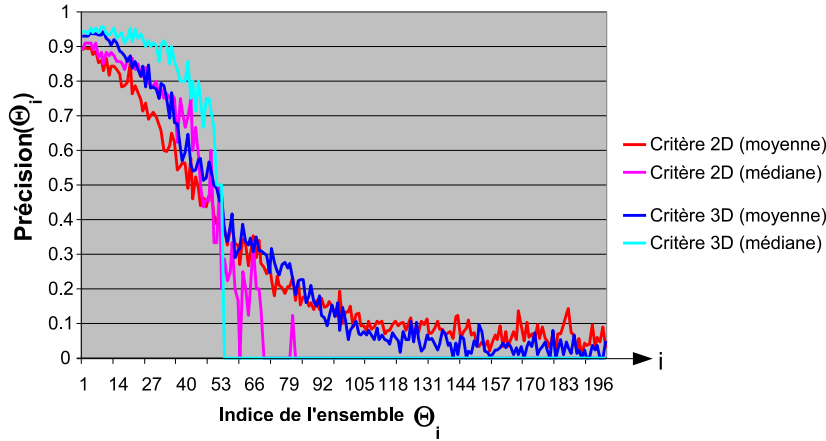


FIG. 5.2 – Première expérience : Moyenne et médiane de $\text{Précision}(\Theta_i)$ calculées sur les 50 vues de la portière.

Pour l'algorithme de filtrage, seule la précision des premiers ensembles Θ importe puisque ce sont à partir de ces seuls ensembles que sont sélectionnés les appariements. Par conséquent, si on compare les résultats obtenus par les différents critères de cohérence, on constate que le critère 3D offre une précision légèrement meilleure que celle du critère 2D. Ce résultat confirme que le critère 3D est un critère plus sévère et donc plus discriminant.

Les résultats obtenus sur l'étalon (figure 5.5) confirment cette observation alors que les deux critères présentent des performances relativement similaires dans le cas de la culasse (figure 5.4).

5.2.3 Le co-rappel

Afin de pouvoir évaluer la robustesse aux poses erronées de chacun des critères de cohérence, nous proposons d'introduire la mesure de **co-rappel**. Celle-ci mesure la quantité de faux appariements contenue dans chaque ensemble Θ_i :

$$\text{Co-Rappel}(\Theta_i) = \frac{\text{Card}(\Theta_i) - \text{Card}(T_{\Theta_i})}{\text{Card}(T_{\Psi})} \quad (5.3)$$

Cette mesure correspond à une normalisation du nombre de faux appariements contenus dans un ensemble Θ_i par le nombre d'appariements corrects contenus

dans Ψ . La quantité de faux appariements sélectionnés est donc exprimée en terme de fraction du nombre maximal d'appariements corrects pouvant être identifié.

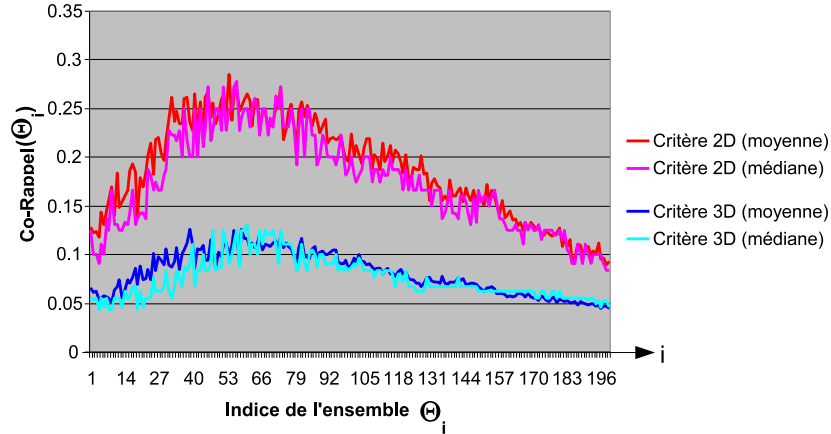
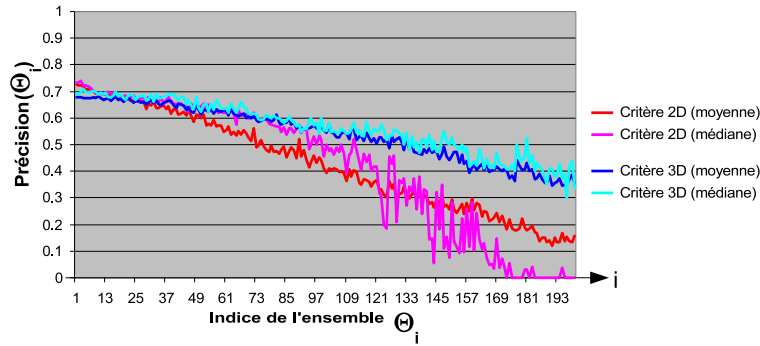


FIG. 5.3 – Première expérience : Moyenne et médiane de $\text{CoRappel}(\Theta_i)$ calculées sur les 50 vues de la portière.

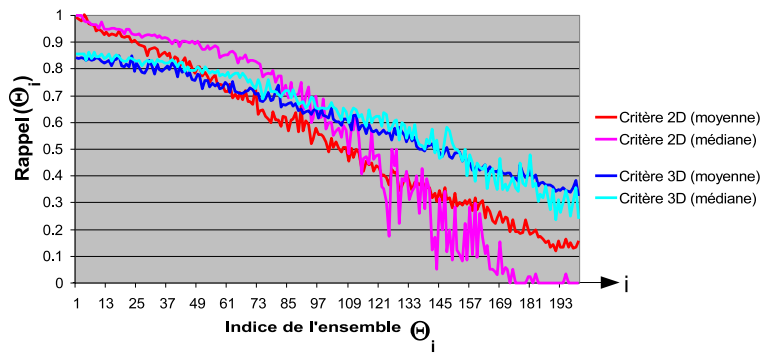
La figure 5.3 présente la mesure du co-rappel obtenue pour la portière. Ces résultats mettent en évidence le fait que le critère 2D sélectionne un plus grand nombre de faux appariements que le critère 3D. Ces résultats confirment une fois de plus le plus grand pouvoir discriminant du critère 3D vis à vis du critère 2D.

De plus, on peut noter que le nombre de faux appariements sélectionnés par le critère 3D varie peu. Cela signifie que la quantité de faux appariements sélectionnés est peu influencée par la qualité des poses \mathbf{P}_i pour lesquelles il est estimé. Le fait que ce critère n'introduise pas plus de faux appariements lorsqu'il est appliqué à une pose erronée que lorsqu'il est appliqué à une pose correcte permet d'assurer au processus de filtrage géométrique une bonne robustesse aux poses erronées.

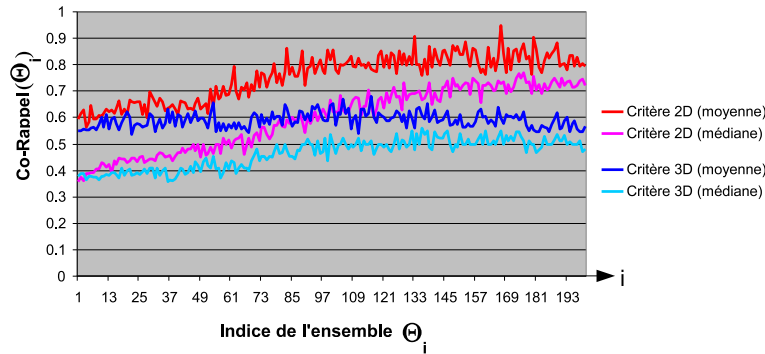
A l'opposé, le critère 2D présente des variations du co-rappel nettement plus importantes. Dans les expériences menées sur la culasse (cf. figure 5.4) et l'étalon (cf. figure 5.5), on observe que le nombre de faux appariements augmente de manière significative lorsque la qualité des poses \mathbf{P}_i décroît. Pour le critère 2D, la quantité de faux appariements sélectionnés dépend donc de la qualité de la pose. Le critère 2D est donc peu robuste aux poses erronées.



(a) Précision

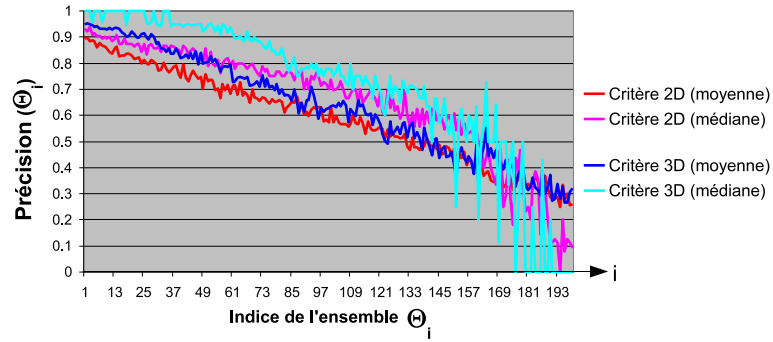


(b) Rappel

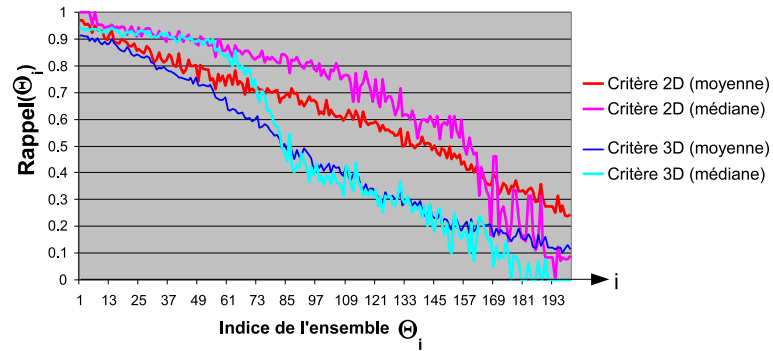


(c) Co-Rappel

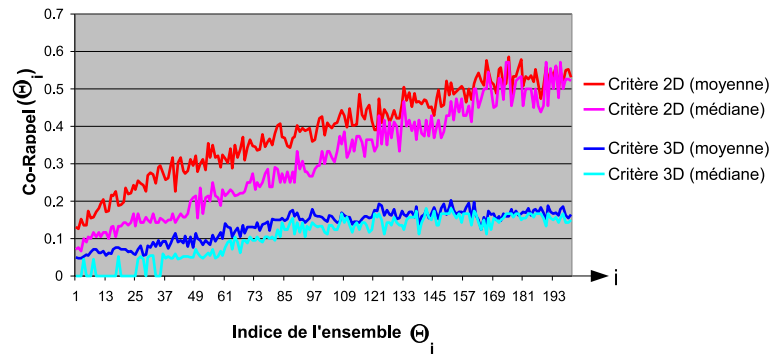
FIG. 5.4 – Résultats de la première expérience pour la Culasse. La première courbe présente la précision obtenue pour chaque ensemble Θ_i , ces derniers étant ordonnés selon la qualité des poses auxquelles ils sont associés. De la même manière, les deux courbes suivantes présentent le rappel et le co-rappel obtenus pour chaque ensemble Θ_i .



(a) Précision



(b) Rappel



(c) Co-Rappel

FIG. 5.5 – Résultats de la première expérience pour l'étalon. La première courbe présente la précision obtenue pour chaque ensemble Θ_i , ces derniers étant ordonnés selon la qualité des poses auxquelles ils sont associés. De la même manière, les deux courbes suivantes présentent le rappel et le co-rappel obtenus pour chaque ensemble Θ_i .

5.3 Performances du processus de filtrage géométrique

Les performances globales du processus de filtrage dépendent du nombre N de poses considérées pour filtrer les appariements et établir l'ensemble d'appariements Θ . Cette expérience consiste à mesurer la précision et le rappel du processus pour un nombre croissant de poses.

Si on note

$$\Xi_j = \bigcup_{i=0..j} \Theta_i$$

alors l'expérience consiste à évaluer $Rappel(\Xi_j)$ et $Précision(\Xi_j)$ avec j allant de 0 à $Card(\Psi)$.

Afin de distinguer les effets du choix du critère de cohérence des effets de la pondération, deux mesures de la précision ont été effectuées :

- une première mesure où tous les appariements de Ξ_j ont le même poids, soit :

$$\forall i, P(< I_i, M_i >) = \frac{1}{Card(\Xi_j)}$$

- une seconde mesure où le poids d'un appariement est défini par l'équation 5.2.

Pour un objet donné, ces mesures sont effectuées pour chacune des images de la sphère de vue de l'objet. Les résultats présentés correspondent aux valeurs moyennes et médianes de ces mesures. Pour améliorer la lisibilité des courbes de résultats, seuls les résultats obtenus pour j allant de 0 à 200 sont présentés.

5.3.1 Le rappel.

La figure 5.6 représente le rappel obtenu lors de l'expérience sur la portière.

Cette seconde expérience confirme que le critère 2D offre de meilleures performances en terme de rappel. Néanmoins, on peut noter que cette différence se réduit dès lors que l'on considère non plus le meilleur ensemble d'appariements Θ_i mais les quelques meilleurs (e.g. les 15 premiers Θ_i).

5.3.2 La précision en l'absence de pondération.

La figure 5.7 présente les résultats obtenus pour la portière. D'une manière générale, on observe une décroissance rapide de la précision, celle-ci pouvant atteindre un score inférieur à 0.5.

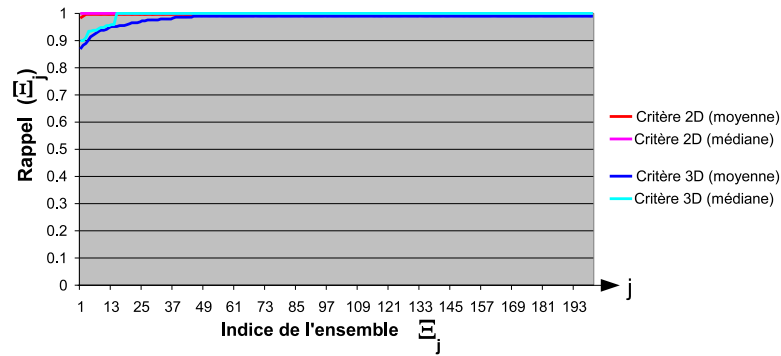


FIG. 5.6 – Seconde expérience : moyenne et médiane de $Rappel(\Xi_j)$ calculées sur les 50 vues de la portière. j représente le nombre de poses utilisées pour établir Ξ_j .

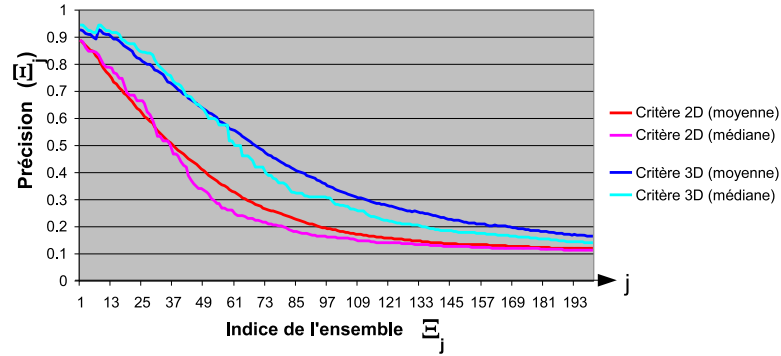


FIG. 5.7 – Seconde expérience : moyenne et médiane de $Rappel(\Xi_j)$ calculées sur les 50 vues de la portière en l'absence de pondération des appariements. j représente le nombre de poses utilisées pour établir Ξ_j .

Si on s'intéresse à l'influence du choix du critère sur le score de précision, on constate que cette seconde expérience confirme les résultats précédents : le critère 3D offre un meilleur score de précision que le critère 2D.

De cette expérience, on peut conclure que l'utilisation des 15 meilleures poses suffisent à identifier plus de 85% des appariements corrects, ceci quelque soit l'objet et le critère utilisé.

5.3.3 La précision avec pondération.

Comme l'illustre la figure 5.8 dans le cas de la portière, la prise en compte de la pondération permet une amélioration notable de la précision du processus

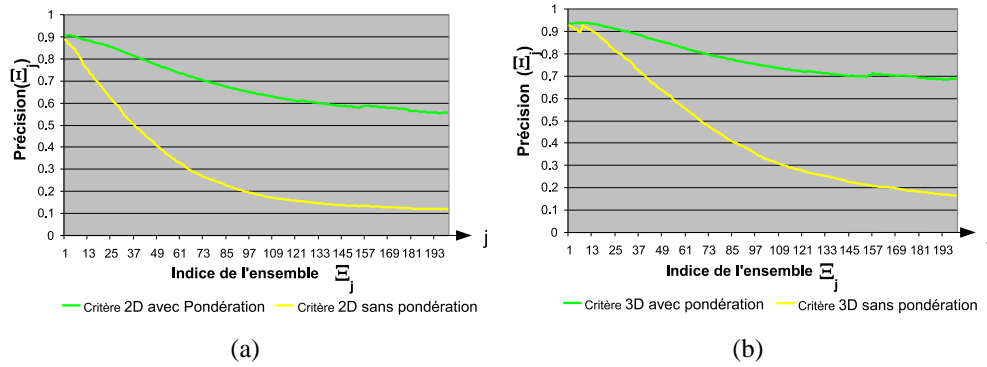


FIG. 5.8 – Influence de la pondération sur la précision : *Précision moyenne, calculées sur les 50 vues la portières, du processus de filtrage géométrique en fonction de la prise en compte ou non de la pondération des appariements, ceci dans le cas du critère 2D (a) et du critère 3D (b).*

de filtrage géométrique, ceci tant pour le critère 2D que 3D. Alors qu'en l'absence de la pondération, la précision décroît rapidement pour descendre sous une précision inférieure à 0.5, la pondération permet de maintenir une précision supérieure à ce seuil en dépit d'une forte proportion de faux appariements. Ainsi, avec près de 80% de faux appariements parmi les appariements sélectionnés (soit une précision sans pondération de 0.2), la probabilité de sélectionner au hasard un appariement correct varie entre 60% et 70% en fonction du critère de cohérence retenu (soit une précision avec pondération entre 0.6 et 0.7). La pondération permet donc de réduire de manière significative l'influence des données aberrantes issues du processus filtrage géométrique.

Si on compare les résultats obtenus selon le critère de cohérence utilisé (cf.

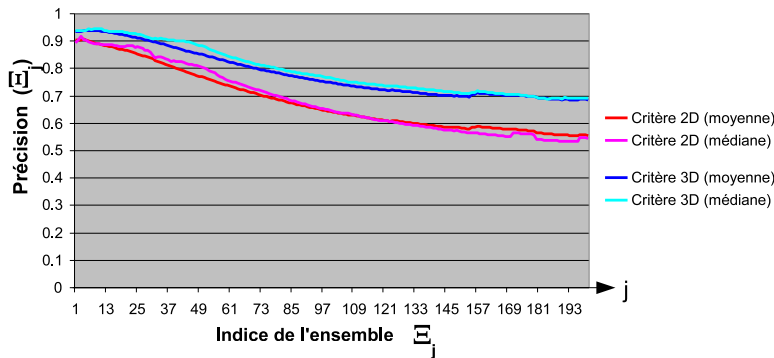
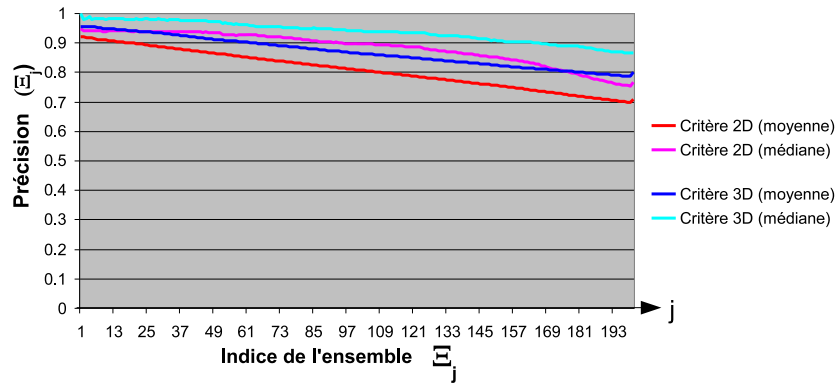


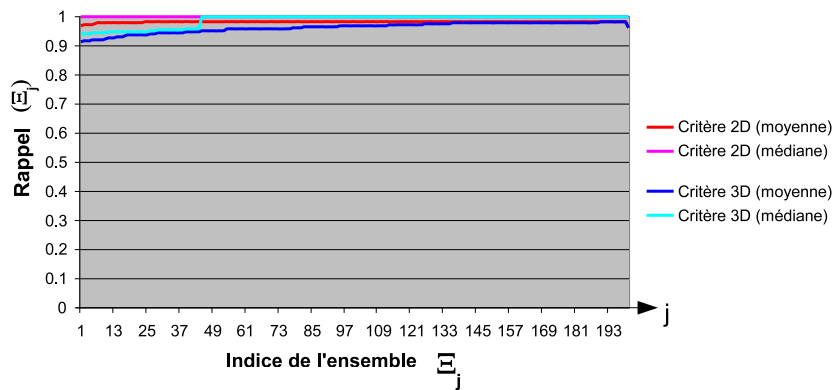
FIG. 5.9 – Seconde expérience : moyenne et médiane de $Precision(\Xi_j)$ calculées sur les 50 vues de la portière avec une prise en compte des poids de appariements.

figure 5.9), on observe toujours un léger avantage pour le critère 3D, celui-ci étant toutefois réduit sous l'effet de la pondération.

Les résultats obtenus pour la culasse et l'étalon sont relativement similaires à ceux obtenus pour la portière, comme le montrent les figures 5.11 et 5.10. De cette expérience, on peut conclure que l'utilisation des 10 premières poses lors du processus de filtrage permet d'établir, si l'on considère la pondération, un ensemble d'appariements comportant plus de 75% d'appariements corrects. En rapportant cette précision à la valeur de rappel obtenu avec le même nombre de poses, constate que le processus de filtrage permet de réduire la proportion de faux appariement à moins de 25% de l'ensemble des appariements filtrés tout en conservant plus de 85% des appariements corrects initiaux, ceci quelque soit l'objet et le critère utilisé.

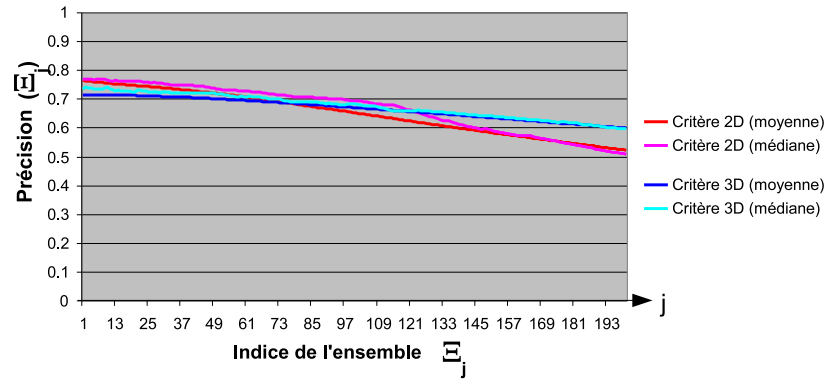


(a) Précision

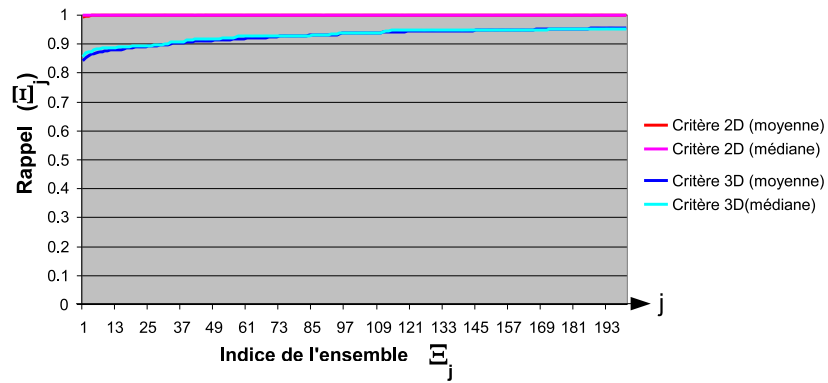


(b) Rappel

FIG. 5.10 – Résultats de la seconde expérience pour l'étalon. *La première courbe présente la précision obtenue pour chaque ensemble Ξ_j , ces derniers étant ordonnés selon le nombre j de poses utilisées pour établir l'ensemble Ξ_j . De la même manière, la seconde courbe présente le rappel obtenu pour chaque ensemble Ξ_j .*



(a) Précision



(b) Rappel

FIG. 5.11 – Résultats de la seconde expérience pour la culasse. La première courbe présente la précision obtenue pour chaque ensemble Ξ_j , ces derniers étant ordonnés selon le nombre j de poses utilisées pour établir l'ensemble Ξ_j . De la même manière, la seconde courbe présente le rappel obtenu pour chaque ensemble Ξ_j .

5.4 Conclusion

Cette évaluation a permis de mettre en évidence l'apport du filtrage géométrique des appariements.

En ce qui concerne le choix du critère de cohérence, on peut conclure que le critère 3D est plus discriminant que le critère 2D mais aussi moins stable. En effet, le critère 3D permet d'extraire des ensembles d'appariements comportant moins de faux appariements. Cependant, la pertinence du critère 3D vis à vis d'un appariement dépend du succès de l'estimation de la pose de ce même appariement. Une mauvaise estimation de cette pose pour un appariement correct peut entraîner son rejet pour le critère 3D alors que celui-ci vérifie généralement le critère 2D. On peut donc conclure que le critère 3D est plus sévère et plus sélectif alors que le critère 2D permet d'identifier une plus grande fraction des appariements corrects.

D'autre part, cette évaluation a permis de souligner l'apport du processus de pondération des appariements. Cette pondération permet de réduire de façon notable l'influence des faux appariements, augmentant ainsi la précision du processus. Les différentes expériences ont montré que cette pondération permet de compenser la différence de pouvoir discriminant entre le critère 2D et le critère 3D. En effet, si le critère 3D couplé à la pondération offre toujours une meilleure précision, l'écart avec celle obtenue par le critère 2D est néanmoins réduit.

A l'issue de cette évaluation, il est donc difficile de départager les deux critères 2D et 3D. Néanmoins, cette évaluation a permis de valider le processus de filtrage géométrique, les résultats obtenus permettant d'envisager une estimation de la pose de l'objet en un temps limité, ceci quelque soit le critère de cohérence employé.

Chapitre 6

Évaluation du processus de recalage

Ce dernier chapitre est consacré à l'évaluation expérimentale du processus de recalage complet. Dans un premier temps, la méthode d'évaluation est présentée (section 6.1). Le processus est ensuite évalué dans trois expériences distinctes. La première évalue la robustesse aux changements de points de vue (section 6.2). La seconde expérience évalue la précision du processus (section 6.3) alors que la dernière (section 6.4) illustre la robustesse de celui-ci en présence d'occultations, de variations d'illumination...

6.1 Protocole expérimental

6.1.1 La base d'objets

Trois objets ont été utilisés pour évaluer le processus de recalage : une culasse de moteur automobile, une portière de voiture ainsi qu'un objet étalon. Pour chacun de ces objets, on dispose d'un modèle 3D ainsi que d'une collection de vues. Pour chacune de ces vues, la vérité terrain, c'est à dire la pose de l'objet vis à vis de la caméra, est parfaitement connue.

Dans le cas de l'étalon et de la portière, les modèles 3D sont de véritables modèles CAO (format IGES) qui nous ont été fournis par la société de métrologie ActiCM. Dans le cas de la culasse, le modèle 3D correspond à une reconstruction par lumière structurée. Il s'agit donc d'un maillage triangulaire et non d'un véritable modèle CAO. Cependant, le processus de recalage n'exploitant pas directement les primitives du modèle CAO, il n'y a pas de contraintes sur la nature de celui-ci. Pour qu'un modèle CAO soit exploitable, il suffit de pouvoir déterminer la position 3D d'un pixel d'une image par un lancer de rayon (cf. section 2.2).

Concernant la vérité terrain, les poses de la portière et de l'étalon ont été fournies par la société de métrologie ActiCM. Les vues ayant été acquises à partir d'une caméra fixée sur un bras robotisé, leurs poses ont été déterminées à partir de la position de ce bras, puis raffinées par un ajustement de faisceaux exploitant des cibles codées placées sur l'objet ainsi que des entités naturelles de celui-ci. Lors des expériences présentées dans ce chapitre, ces cibles codées ne sont pas exploités par le processus de recalage.

Pour la culasse, les poses ont été estimées par nos soins à l'aide d'une mire plane¹ dont la géométrie est connue. Ces poses ont été raffinées par un ajustement de faisceaux intégrant des primitives naturelles de l'objet (segments de droites). Enfin, quelque soit l'objet considéré, les paramètres intrinsèques de la caméra ont été déterminés à l'aide d'une mire de calibrage 3D.

Préalablement au recalage, le modèle de l'objet a été construit à partir du modèle CAO et de N vues de l'objet (N allant de 4 à 6 en fonction de l'objet considéré). Dans toutes les expériences, le modèle est élaboré avec des descripteurs STHGI comportant 5 anneaux et 4 secteurs ainsi que sur des descripteurs *Shape Context* comportant 3 anneaux et 64 secteurs.

La culasse de moteur

- Dimension : $390mm \times 200mm \times 170mm$
- Collection de vues : 80 vues (1280×960 pixels) prises à une distance de 1000mm (cf. figure 6.1)
- Ecart angulaire maximal entre deux points de vue : 125°
- focale de la caméra : 12 mm
- Eclairage : contrôlé (salle noire avec une éclairage annulaire embarqué sur la caméra)

La portière

- Dimension : $1260mm \times 1050mm \times 120mm$
- Collection de vues : 50 vues (1392×1040 pixels) prises à une distance de 1500mm (cf. figure 6.2).
- Ecart angulaire maximal entre deux points de vue : 95°
- focale de la caméra : 8 mm
- Eclairage : non contrôlé

¹Lors des expériences, la mire n'est pas exploités par le processus de recalage.

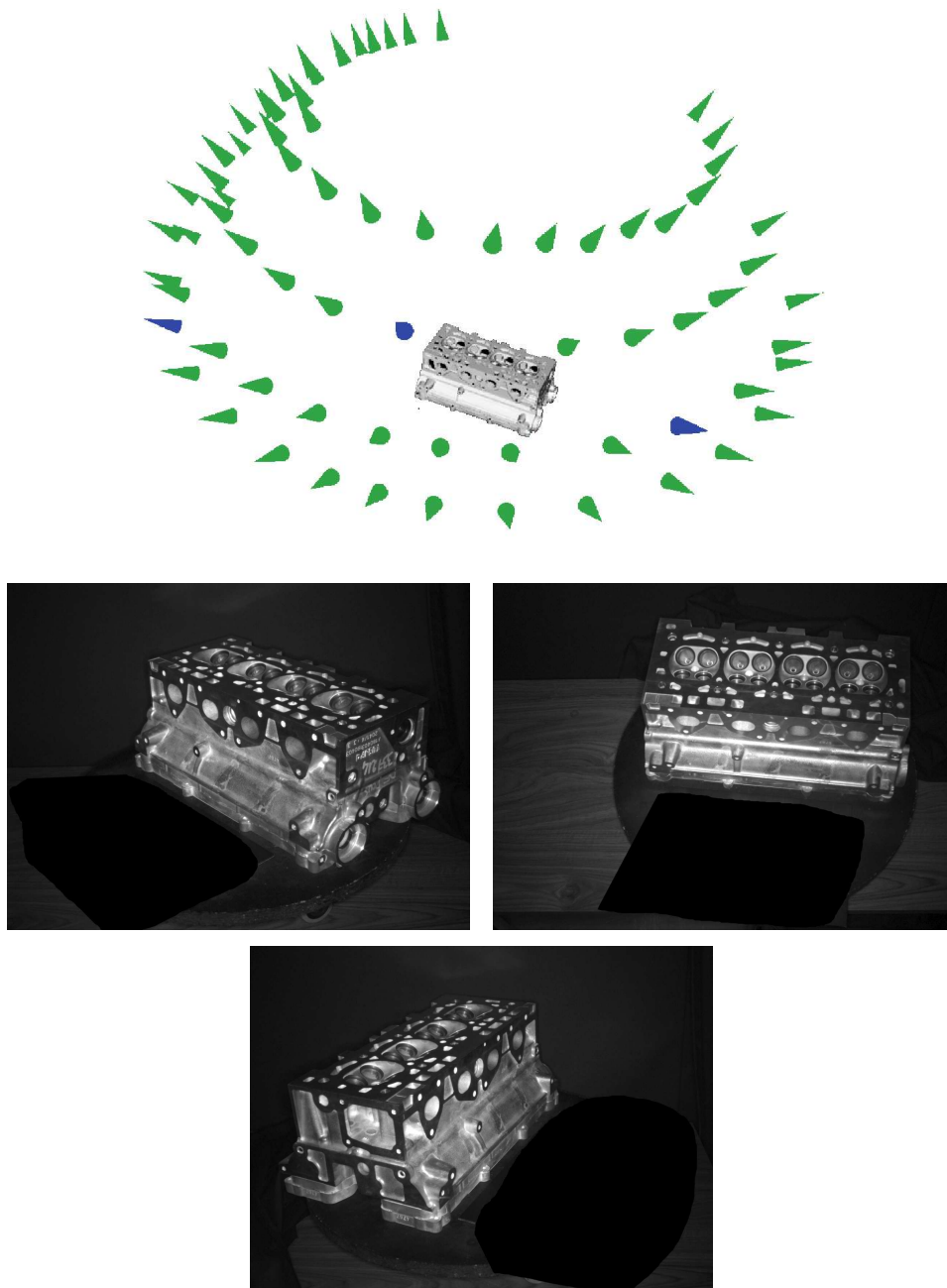


FIG. 6.1 – La culasse de moteur automobile. La première illustration représente le modèle CAO et la sphère de vues. Les cônes représentent les différents points de vue formant la sphère de vues. Les trois images suivantes correspondent aux vues utilisées pour élaborer le modèle, les points de vue correspondants sont représentés par des cônes bleus dans la première illustration.

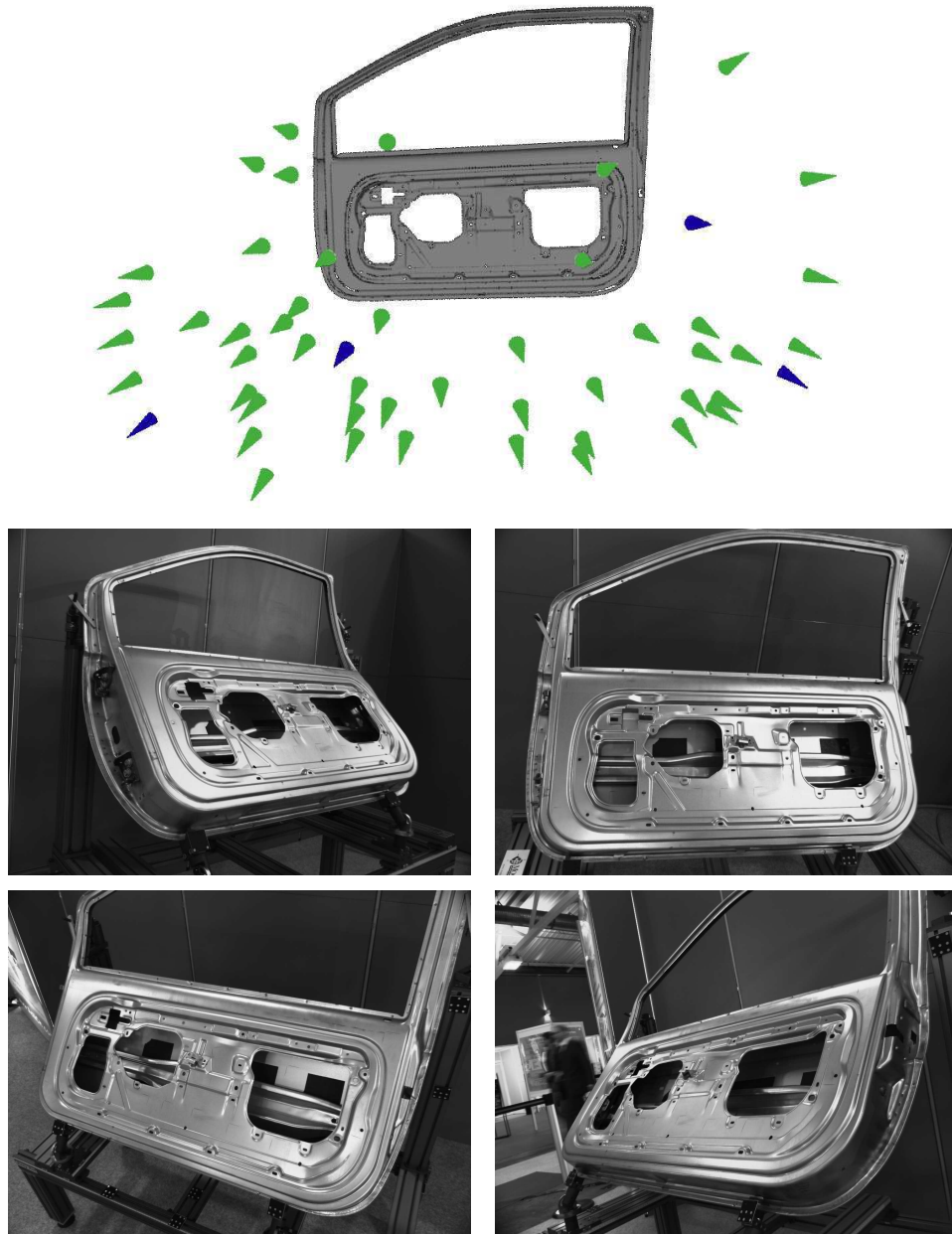


FIG. 6.2 – La portière automobile. La première illustration représente le modèle CAO et la sphère de vue. Les cônes représentent les différents points de vue formant la sphère de vues. Les quatre images correspondent aux vues utilisées pour élaborer le modèle, les points de vue correspondants sont représentés par des cônes bleus dans la première illustration.

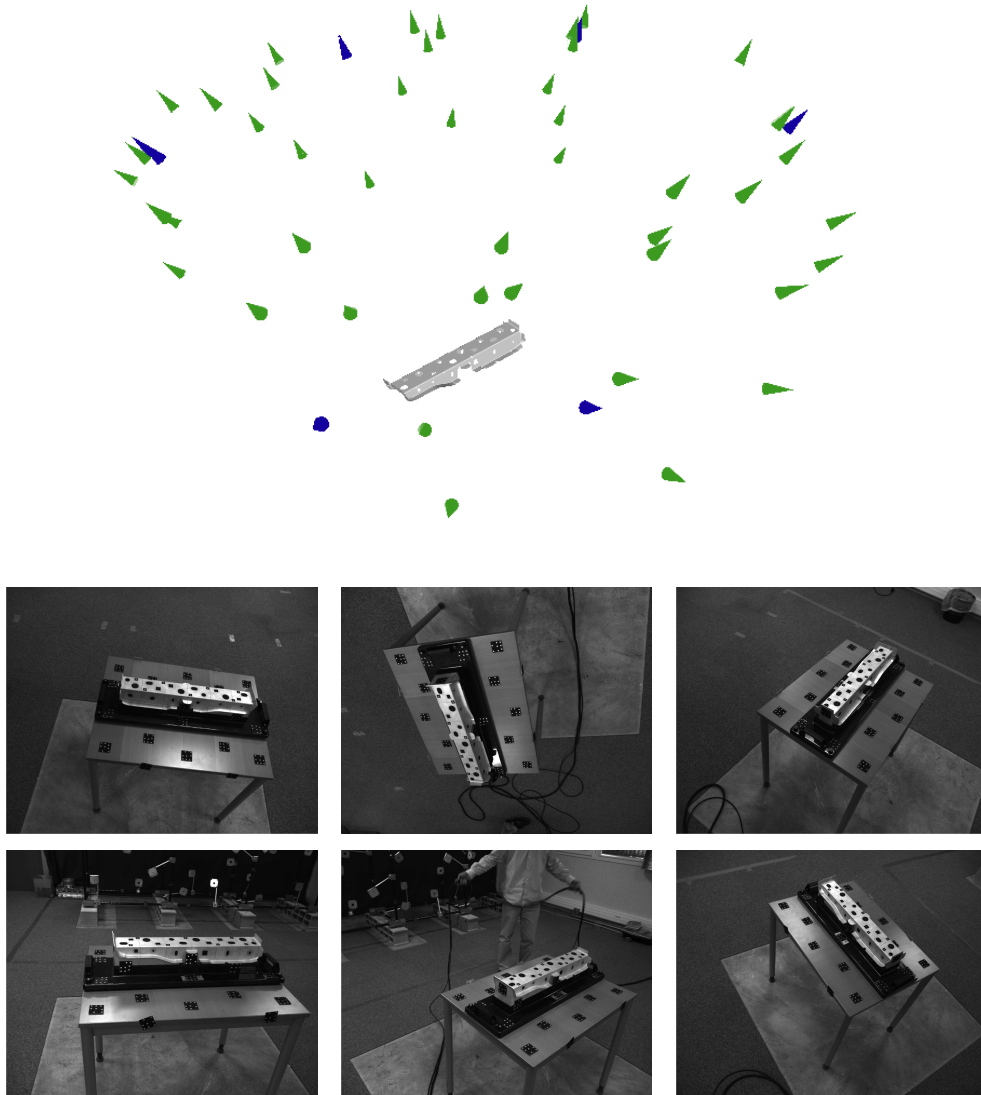


FIG. 6.3 – L'étalon métallique. La première illustration représente le modèle CAO et la sphère de vues. Les cônes représentent les différents points de vue formant la sphère de vue de l'étalon. Les six images correspondent aux vues utilisées pour élaborer le modèle, les points de vue correspondants sont représentés par des cônes bleus dans la première illustration.

L'étalon

- Dimension : $600mm \times 90mm \times 130mm$
- Collection de vues : 50 vues (1300×1030 pixels) prises à une distance de 1400mm (cf. figure 6.3).
- Ecart angulaire maximal entre deux points de vue : 140°
- Eclairage : non contrôlé

6.1.2 Caractérisation de la qualité du recalage

Trois mesures sont proposées pour évaluer la qualité d'un recalage. La première correspond à l'erreur d'estimation des paramètres de la pose. La seconde correspond à l'erreur de reprojection des points de surface de l'objet. La troisième correspond à l'erreur de position 3D des points de surface de l'objet.

Erreur sur les paramètres de la pose

Cette mesure représente l'erreur commise sur l'estimation de la pose. La vérité terrain étant connue pour chaque image de la base, cette erreur correspond à la transformation rigide reliant la pose estimée à la pose réelle de l'objet. Cette erreur comprend :

- l'erreur de position : il s'agit de la norme (exprimée en mm) du vecteur translation de la transformation rigide ;
- l'erreur d'orientation : il s'agit de l'angle (en radian) de rotation de la transformation rigide.

La transformation rigide est estimée au niveau du centre de gravité de l'objet étudié.

Erreur de reprojection des points de surface

Cette erreur mesure l'écart entre la projection d'un point de surface du modèle et sa position réelle dans l'image. Notons \mathbf{P}_{Reelle} la pose réelle de l'objet (la vérité terrain) et P_{Estim} la pose estimée par l'algorithme. A partir de la pose \mathbf{P}_{Reelle} , il est possible de déterminer le point de surface de l'objet $\mathbf{S}_{x,y} = (X, Y, Z)^T$ se projetant dans l'image sur le pixel (x, y) . L'erreur de reprojection Δ_{proj} associée au pixel (x, y) correspond alors à la distance euclidienne entre ce point (x, y) et la projection de $\mathbf{S}_{x,y}$ selon la pose P_{Estim} :

$$\Delta_{proj}(x, y) = \|\mathbf{K} \cdot \mathbf{P}_{Reelle} \cdot \mathbf{S}_{x,y} - \mathbf{K} \cdot \mathbf{P}_{Estim} \cdot \mathbf{S}_{x,y}\|$$

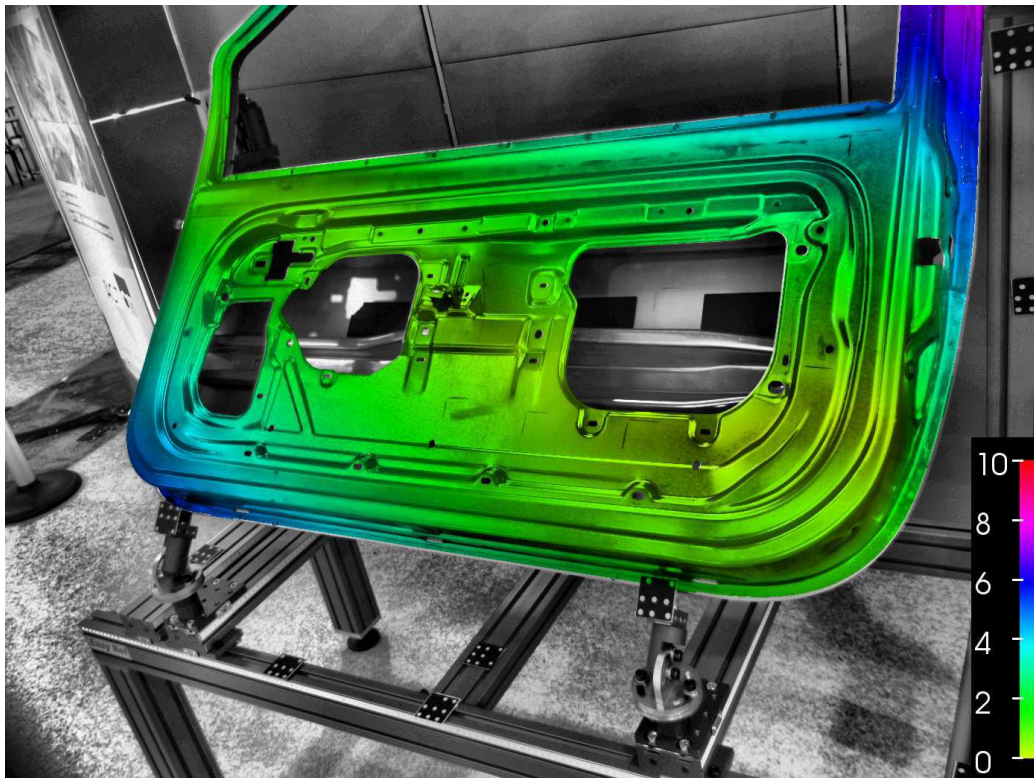


FIG. 6.4 – Représentation en fausses couleurs de l'erreur de reprojection. *La couleur des pixels reflète l'erreur de reprojection. L'échelle à droite représente la couleur du pixel en fonction de l'erreur de reprojection (exprimée en pixels). Cette représentation permet de visualiser les variations de l'erreur de reprojection.*

L'erreur Δ_{Proj} peut donc être évaluée pour chaque pixel de l'image appartenant à l'objet. L'erreur de reprojection n'est donc pas mesurée uniquement au niveau des appariements utilisés pour estimer la pose mais sur l'ensemble de la surface de l'objet visible à l'image. Cette erreur n'étant pas nécessairement uniforme à la surface de l'objet, il est possible de représenter les variations de l'erreur de reprojection par une image en fausses couleurs comme illustrée dans la figure 6.4. Une mesure globale sur l'ensemble de la surface de l'objet peut aussi être obtenue en calculant la moyenne des erreurs.

Erreur sur la position 3D des points de surface

Cette erreur consiste à mesurer, pour chaque point 3D de surface visible, la distance entre sa position réelle et sa position estimée. Afin de pouvoir représenter

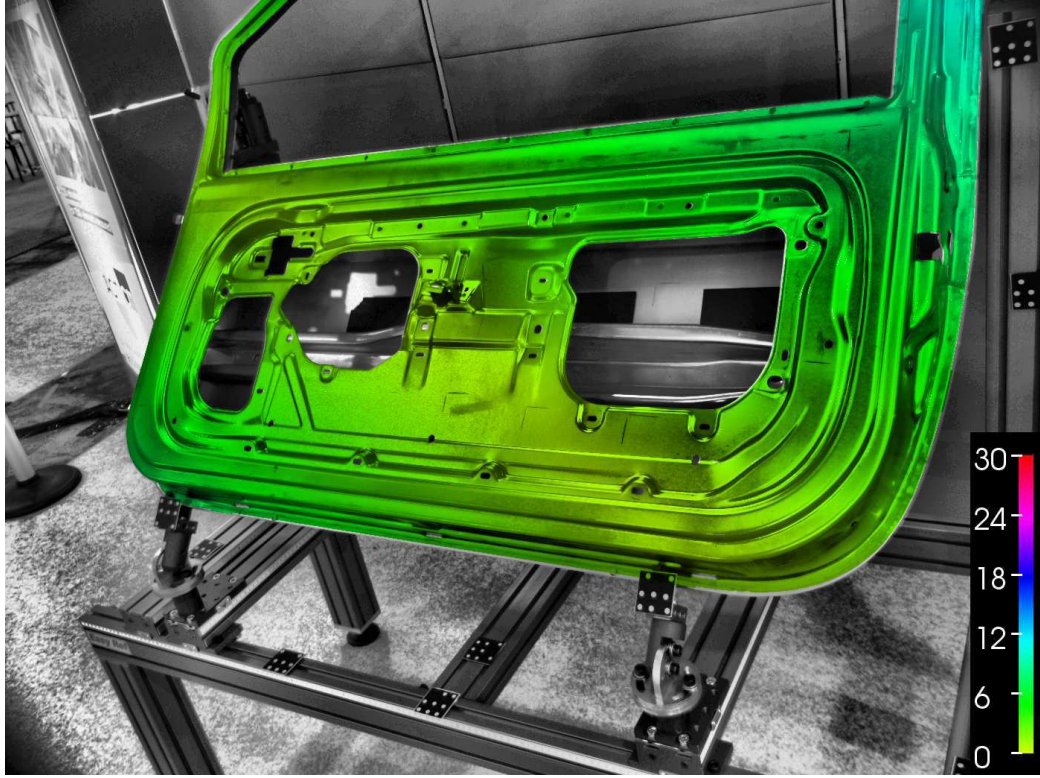


FIG. 6.5 – Représentation en fausses couleurs de l'erreur de position. La couleur des pixels reflète l'erreur de position commise sur le point de surface de l'objet correspondant. L'échelle à droite représente la couleur du pixel en fonction de l'erreur de position (exprimée en mm). Cette représentation permet de visualiser les variations de l'erreur de position à la surface de l'objet.

la distribution de cette erreur à la surface de l'objet, celle-ci est mesurée pour chacun des points de surface de l'objet se projetant à l'image. Ainsi, si on note $\mathbf{S}_{x,y}$ le point de surface de l'objet se projetant à l'image sur le pixel (x,y) , l'erreur de position Δ_{Pos} associée au pixel (x,y) est définie par :

$$\Delta_{Pos}(x,y) = \|\mathbf{P}_{Reelle} \cdot \mathbf{S}_{x,y} - \mathbf{P}_{Estim} \cdot \mathbf{S}_{x,y}\|$$

où \mathbf{P}_{Reelle} correspond à la pose réelle de l'objet (la vérité terrain), \mathbf{P}_{Estim} correspond à la pose estimée par l'algorithme et $\|\cdot\|$ correspond à la norme euclidienne.

L'erreur Δ_{Pos} peut donc être évaluée pour chaque pixel de l'image appartenant à l'objet. Par conséquent, l'erreur de position n'est pas mesurée uniquement au niveau des appariements utilisés pour estimer la pose mais sur l'ensemble de la

surface de l'objet visible à l'image. Cette erreur n'étant pas nécessairement uniforme à la surface de l'objet, il est possible de représenter ses variations par une image en fausses couleurs comme illustrée dans la figure 6.5. Une mesure globale sur l'ensemble de la surface de l'objet peut aussi être obtenue en calculant la moyenne des erreurs.

6.1.3 Algorithmes évalués

L'algorithme de recalage visuel proposé peut se décliner en plusieurs versions en fonction du critère de cohérence des appariements utilisé. En effet, comme exposé dans la section 4.1.3, deux critères différents (*ErrProj* et *Err3D*) permettent d'évaluer la cohérence d'un appariement pour une hypothèse de pose. Le premier correspond à un critère 2D (une erreur de reprojection) alors que le second correspond à un critère 3D (erreur des trièdres). Ce critère intervenant à la fois lors de l'étape de la mise en correspondance (cf. section 4.1.3) et lors du calcul de la pose (cf. section 4.2.1), l'algorithme peut se décliner sous 4 formes. Parmi celles-ci, trois ont été évaluées et sont désignées de la manière suivante :

- le processus 2D/2D : le critère 2D est employé à la fois lors de la mise en correspondance et du calcul de pose ;
- le processus 3D/2D : le critère 3D est employé lors de la mise en correspondance puis le critère 2D est employé pour estimer la pose ;
- le processus 3D/3D : le critère 3D est employé à la fois lors de la mise en correspondance et du calcul de pose.

Concernant le seuil sur l'erreur de cohérence, celui-ci est fixé à 70 pixels d'erreur de reprojection pour le critère 2D lors de l'étape de mise en correspondance et à 10 pixels lors de l'étape de calcul de pose. Pour le critère 3D, ce seuil est atteint pour une erreur de translation supérieure ou égale à $1/5^{\text{ème}}$ de la longueur de l'objet ou une erreur d'orientation supérieure ou égale à 70° .

6.2 Robustesse aux changements de points de vue

Cette expérience a pour objectif d'évaluer la robustesse du processus de recalage aux changements de points de vue. L'expérience consiste à élaborer le modèle de l'objet à partir d'une seule vue de l'objet, appelée *vue d'apprentissage*. Ce modèle est ensuite utilisé pour recaler chacune des vues de l'objet et calculer l'erreur de reprojection de celles-ci. Ces erreurs peuvent alors être exprimées en fonction de l'écart angulaire (cf. figure 6.6) entre la vue recalée et la vue d'apprentissage. Cette expérience est ensuite répétée autant de fois que la sphère contient de vues,

le modèle étant à chaque fois élaboré à partir d'une vue d'apprentissage différente.

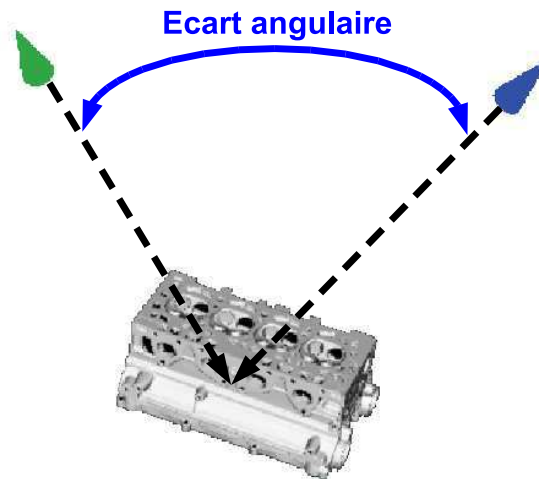


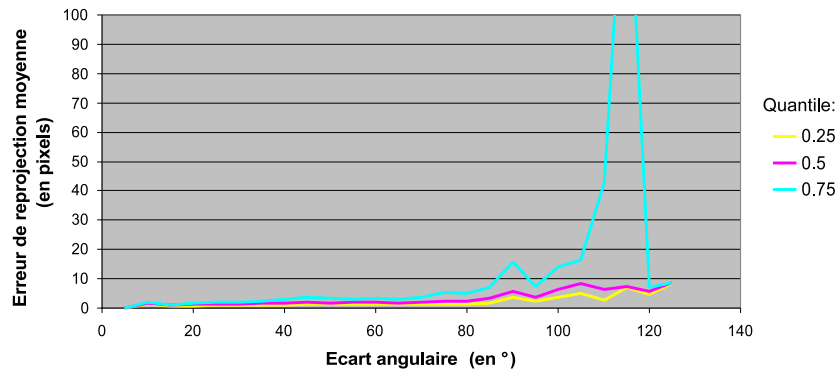
FIG. 6.6 – L'écart angulaire entre deux vues. *Le cône bleu représente la pose de la caméra pour la vue apprise alors que le cône vert représente la pose de la caméra de la vue à recalier.*

Cette expérience a été menée sur les trois objets et pour les trois variantes du processus de recalage. Les résultats obtenus pour la portière, la culasse et l'étalon sont respectivement présentés dans les figures 6.9, 6.7 et 6.8. Pour chaque variante du processus de recalage, les résultats sont présentés sous la forme de trois courbes : la première correspond à l'erreur de reprojection médiane (ou quantile 0.5), calculée sur l'ensemble des expériences, en fonction de l'écart angulaire, alors que les deux autres correspondent au quartile (ou quantile 0.25) et au quantile 0.75 des erreurs de reprojections. La courbe correspondant au quantile 0.75 (resp. 0.5 et 0.25) peut être interprétée comme étant la précision assurée dans 75% (resp. 50% et 25%) des cas.

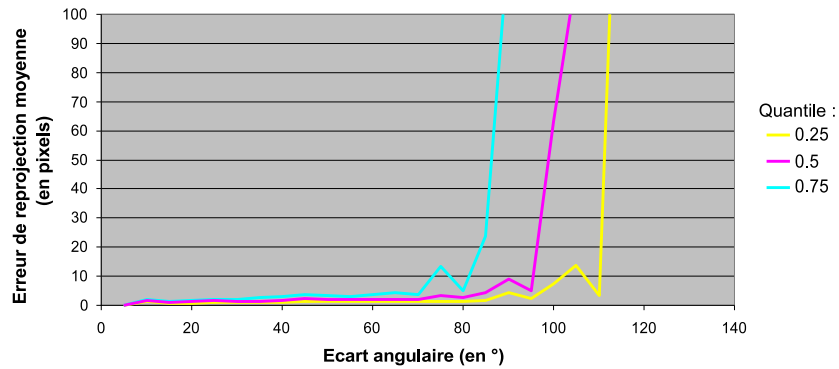
Tout d'abord, on peut observer que la robustesse aux changements de points de vue n'est pas la même en fonction de la pièce étudiée. Ainsi, avec la culasse, et dans le cas de la méthode 2D/2D, on obtient 75% de recalages corrects pour des écarts angulaires allant jusqu'à 80° , alors qu'un tel taux de réussite ne peut être assuré au delà de 20° dans le cas de la portière. Ces différences s'expliquent par la nature des pièces. En particulier, la culasse présente des surfaces plus planes que celles de la portière. L'approximation, par une transformation affine, des transformations induites par un changement de point de vue sur l'apparence d'une région

support, est donc valide pour des changements de points de vue plus importants. De plus, la portière présente des alésages à la fois plus petits à l'image, ce qui nuit à la précision des poses estimées à partir d'un seul appariement, et moins nombreux, ce qui réduit la robustesse du calcul de pose. Ces premiers résultats soulignent donc la nécessité d'adapter l'écart angulaire entre les vues d'apprentissage (et donc leur nombre) en fonction de l'objet considéré. Ces résultats seront pris en compte dans les expériences suivantes lors de l'élaboration du modèle.

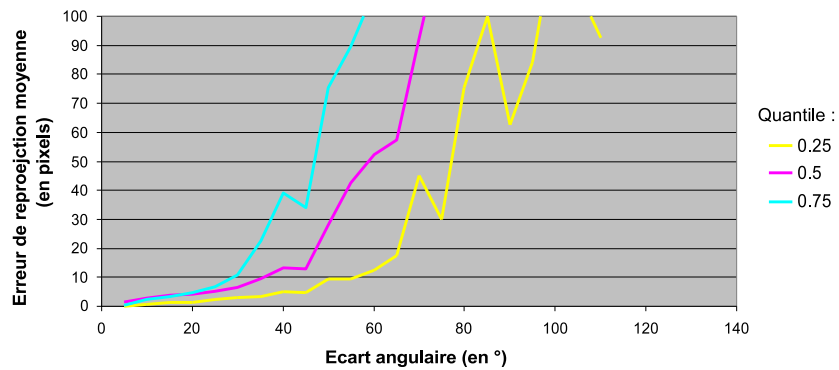
D'autre part, on peut observer que le recalage offrant la meilleure robustesse aux changements de points de vue correspond au recalage 2D/2D. Les performances obtenues par le recalage 3D/2D sont cependant très proches de celles-ci. Concernant les performances du recalage 3D/3D, celles-ci sont inférieures à celles obtenues par les deux précédentes méthodes.



(a) Recalage 2D/2D

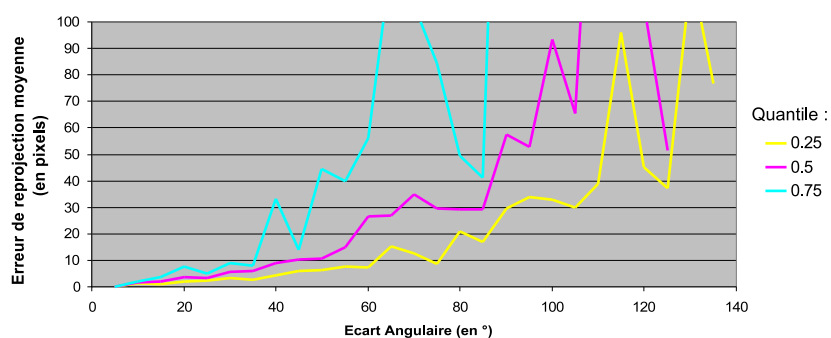


(b) Recalage 3D/2D

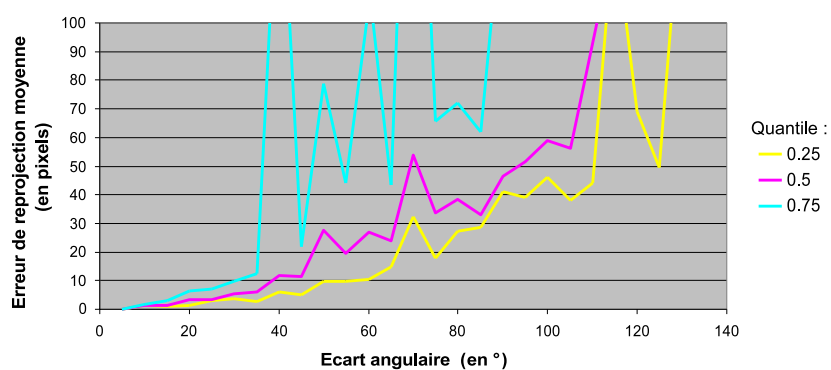


(c) Recalage 3D/3D

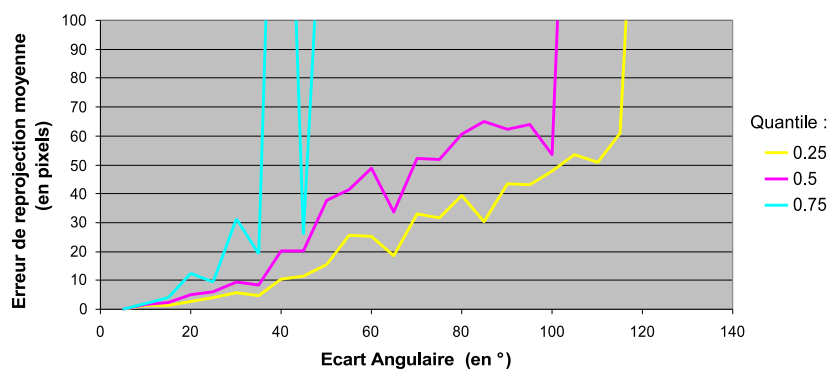
FIG. 6.7 – Evaluation de la robustesse aux changements de points de vue pour la culasse. Ces graphiques représentent, pour différents quantiles (0.25, 0.5 et 0.75), l'évolution de l'erreur de reprojection du modèle en fonction de l'écart angulaire entre la vue recalée et la vue utilisée pour élaborer le modèle. Les résultats présentés dans les graphes (a), (b) et (c) correspondent respectivement aux résultats obtenus avec un recalage 2D/2D, 3D/2D et 3D/3D.



(a) Recalage 2D/2D

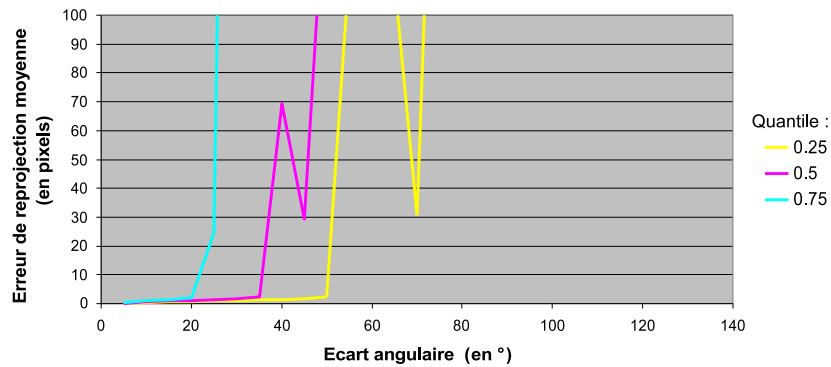


(b) Recalage 3D/2D

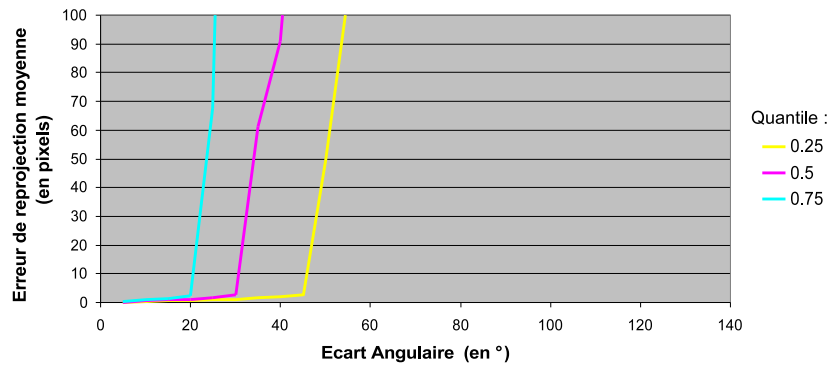


(c) Recalage 3D/3D

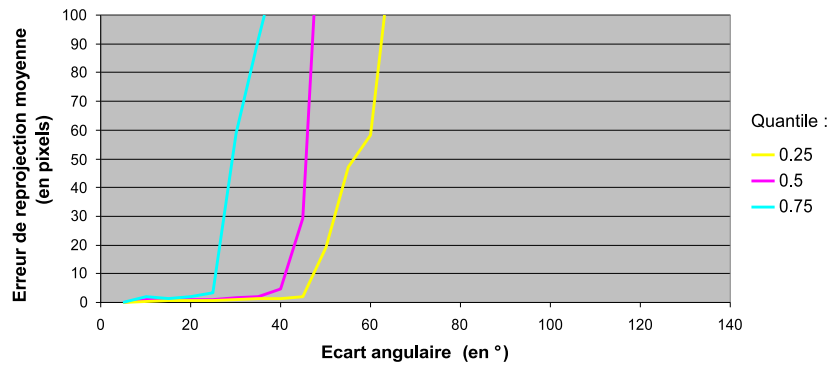
FIG. 6.8 – Évaluation de la robustesse aux changements de points de vue pour l'étalon. Ces graphiques représentent, pour différents quantiles (0.25, 0.5 et 0.75), l'évolution de l'erreur de reprojection du modèle en fonction de l'écart angulaire entre la vue recalée et la vue utilisée pour élaborer le modèle. Les résultats présentés dans les graphes (a), (b) et (c) correspondent respectivement aux résultats obtenus avec un recalage 2D/2D, 3D/2D et 3D/3D.



(a) Recalage 2D/2D



(b) Recalage 3D/2D



(c) Recalage 3D/3D

FIG. 6.9 – Evaluation de la robustesse aux changements de points de vue pour la portière. Ces graphiques représentent, pour différents quantiles (0.25, 0.5 et 0.75), l'évolution de l'erreur de reprojection du modèle en fonction de l'écart angulaire entre la vue recalée et la vue utilisée pour élaborer le modèle. Les résultats présentés dans les graphes (a), (b) et (c) correspondent respectivement aux résultats obtenus avec un recalage 2D/2D, 3D/2D et 3D/3D.

6.3 Performances avec un modèle léger

Cette expérience évalue les performances du processus de recalage dans le cadre d'une application conventionnelle, c'est à dire à l'aide d'un modèle élaboré à partir de quelques images couvrant les différents points de vue. La robustesse du processus aux changements de points de vues permet d'élaborer un tel modèle à partir d'un nombre réduit de vues, modèle alors qualifié de *léger* (à opposer à des modèles *lourds* nécessitant plusieurs centaines de vues pour un objet).

Cette expérience a été menée sur la culasse de moteur, la portière ainsi que sur l'objet étalon. La couverture des sphères de vues et la robustesse du processus aux changements de points de vue variant en fonction de l'objet considéré, le nombre de vues employées pour élaborer le modèle de chaque objet est différent : 3 pour la culasse, 4 pour la portière et 6 pour l'étalon. Ces vues sont représentées par les cônes bleus dans les figures 6.1, 6.2 et 6.3.

6.3.1 Échecs et réussites du recalage

On considère ici qu'un recalage a échoué dès lors que son erreur de reprojexion moyenne dépasse les 10 pixels. On peut donc déterminer le taux d'échec pour chaque objet à partir des courbes de résultats de la figure 6.11.

S'agissant des approches 2D/2D et 3D/2D, on peut noter que celles-ci présentent des taux de succès particulièrement élevés puisque la totalité des vues de la culasse et de la portière sont correctement recalées ainsi que 94% des vues de l'étalon. De plus, les échecs connus dans le cas de l'étalon s'expliquent, pour partie, par la géométrie particulière de la pièce. En effet, les alésages présents sur les faces avant et arrière de cette pièce sont tous alignés le long d'un axe. Or, pour pouvoir estimer une pose à partir d'appariements, il est nécessaire que les points 3D appariés ne soient pas alignés. Dans le cas contraire, la pose ne peut être déterminée qu'à une rotation près autour de cet axe. Par conséquent, lorsque l'étalon est observé selon un point de vue où seule l'une des faces avant ou arrière de l'étalon est visible, la pose ne peut être totalement déterminée. Un tel cas de figure est rencontré à plusieurs reprises au cours de l'expérience, comme l'illustre la figure 6.10.

S'agissant de l'approche 3D/3D, le taux d'échec est nettement plus élevé (près de 25% d'échec pour la culasse). Notons cependant que de nombreux échecs correspondent à des recalages très approximatifs, dépassant largement le seuil des 10 pixels d'erreur de reprojexion. Si l'on considère uniquement la proportion de recalages totalement incohérents avec la pose réelle, c'est à dire ne permettant

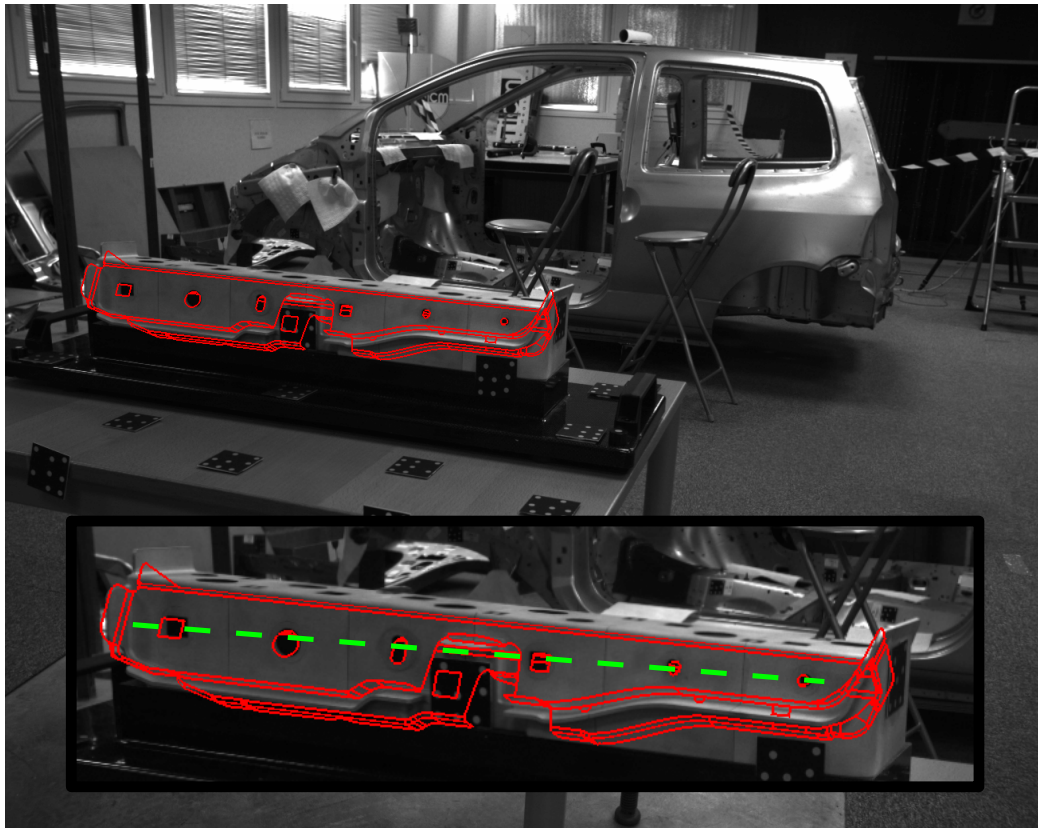


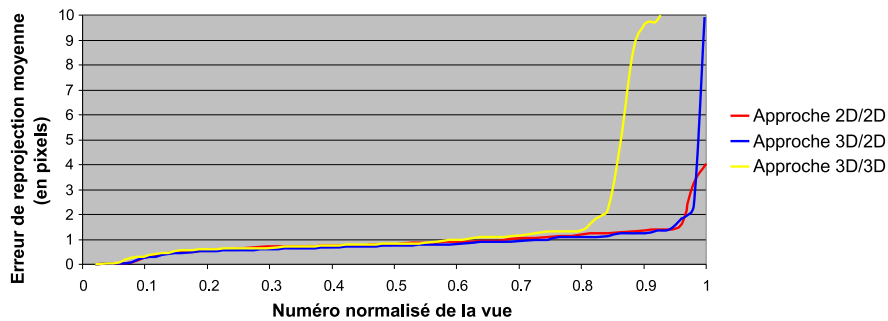
FIG. 6.10 – Configuration ambiguë pour l'étalon. *Les alésages de la seule face visible sont alignés selon un axe symbolisé par des pointillés verts. On observe que l'erreur de recalage correspond à une rotation autour de cet axe.*

pas de localiser l'objet même approximativement, l'étude des images de recalage révèle que ce taux ne dépasse pas les 10% au cours de chaque expérience.

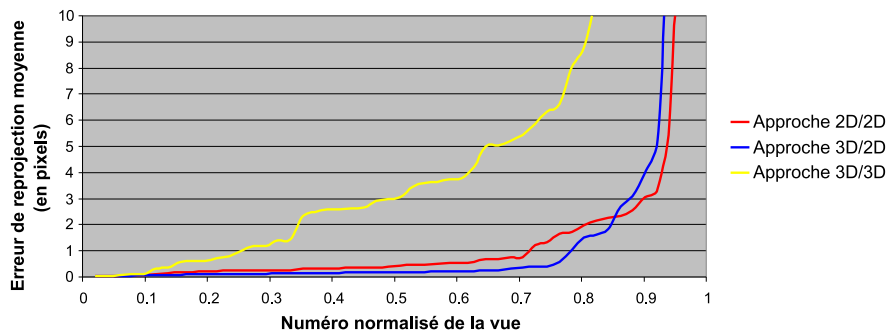
6.3.2 Qualité visuelle du recalage.

Les erreurs de reprojections obtenues pour les différents objets sont présentées dans la figure 6.11. La figure 6.12 permet d'apprécier visuellement la qualité d'un recalage avec des erreurs de reprojection de 1 et 4 pixels.

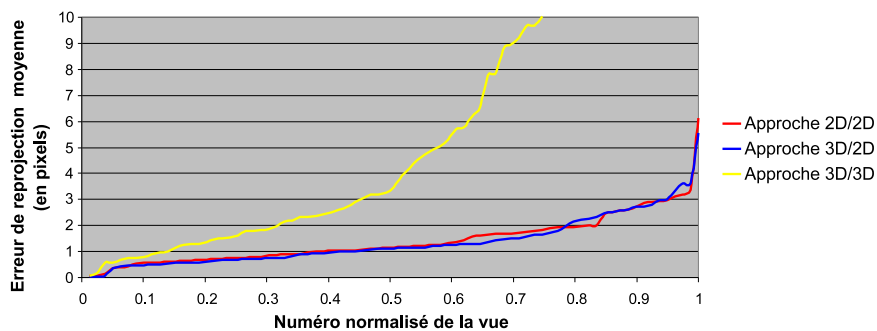
Les approches 2D/2D et 3D/2D offrent une précision visuelle équivalente. Quelque soit l'objet considéré, l'erreur moyenne de reprojection est inférieure à 2 pixels dans plus de 80% des vues. L'approche 3D/3D fournit des recalages plus approximatifs. Celle-ci permet néanmoins d'obtenir des recalages dont l'erreur de



(a) Résultats pour la portière



(b) Résultats pour l'étalon



(c) Résultats pour la culasse

FIG. 6.11 – Qualité visuelle du recalage *Le graphe (a) représente l'erreur de re-projection pour chacune des 50 vues de la portière, celles-ci étant triées par ordre croissant d'erreur. Les vues sont indicées selon une numérotation normalisée : la vue ayant l'erreur minimale correspond au numéro 0, la vue ayant l'erreur maximale correspond au numéro 1 et la vue ayant l'erreur médiane correspond à l'abscisse 0.5. De la même manière, les graphes (b) et (c) représentent les erreurs de re-projections obtenues pour l'étalon et la culasse.*

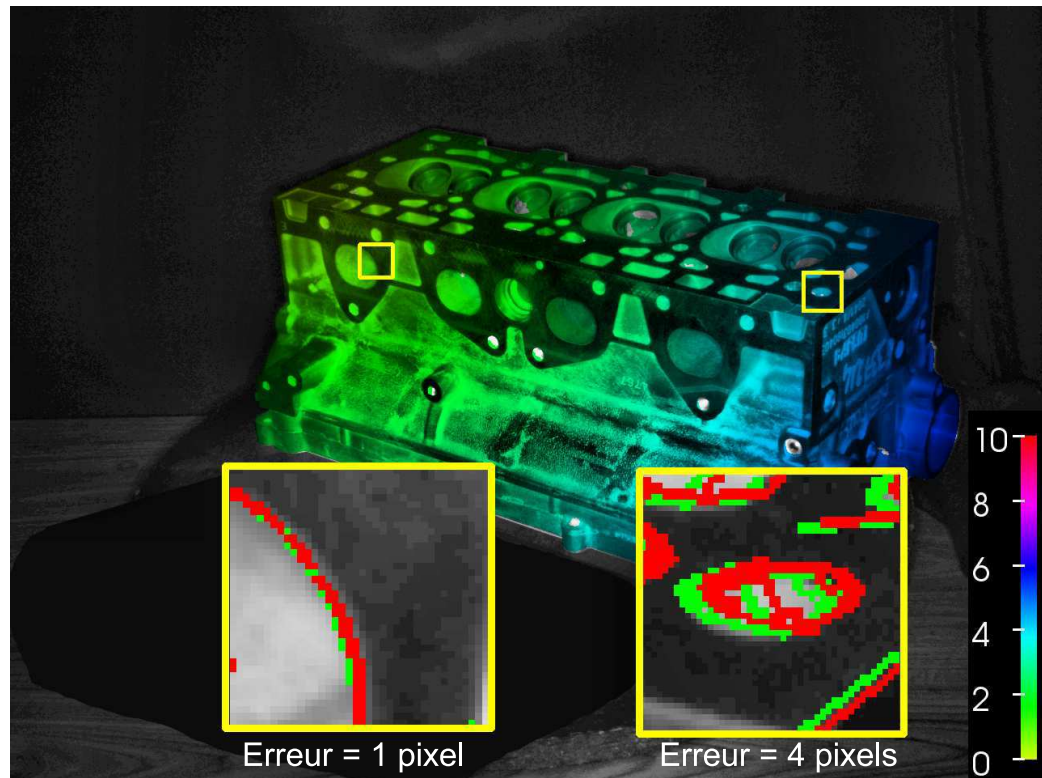


FIG. 6.12 – Exemples d’erreurs de reprojection. L’image de fond représente la distribution de l’erreur de reprojection en fausses couleurs. Les cadres jaunes correspondent à des zooms de deux parties de la culasse. On y voit la reprojection du modèle selon la pose réelle de l’objet (en vert) et celle selon la pose estimée (en rouge).

reprojection est inférieure à 5 pixels dans plus de 60% des vues (cf. figure 6.11).

Pour chacune des trois méthodes, les images-résultats correspondant au meilleur recalage, au pire recalage et à un recalage moyen de la portière, sont présentées dans les figures 6.20, 6.22, 6.24. Les images-résultats du recalage moyen pour les différentes méthodes sont présentées en pleines pages dans les figures 6.27, 6.30 et 6.33. Les résultats obtenus avec la culasse et l'étalon étant relativement similaires, ceux-ci sont présentés dans l'annexe C.

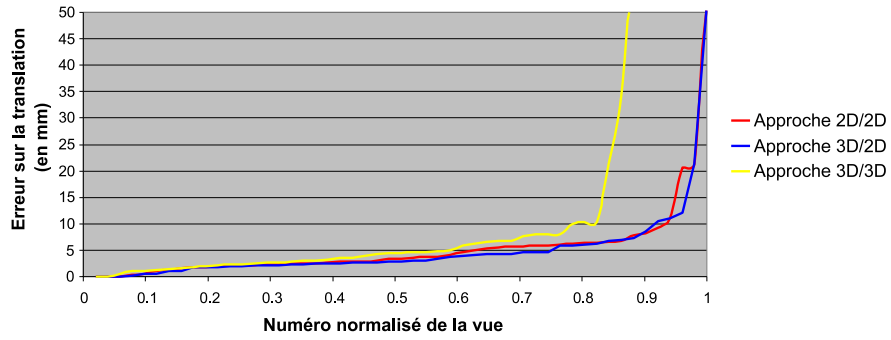
D'autre part, on notera que la répartition des erreurs à la surface de l'objet est généralement relativement uniforme, comme l'illustrent les séries d'imagettes des figures 6.17 à 6.19 pour la portière (les images-résultats obtenus pour l'étalon et la culasse sont présentées dans l'annexe C).

6.3.3 Précision de la pose

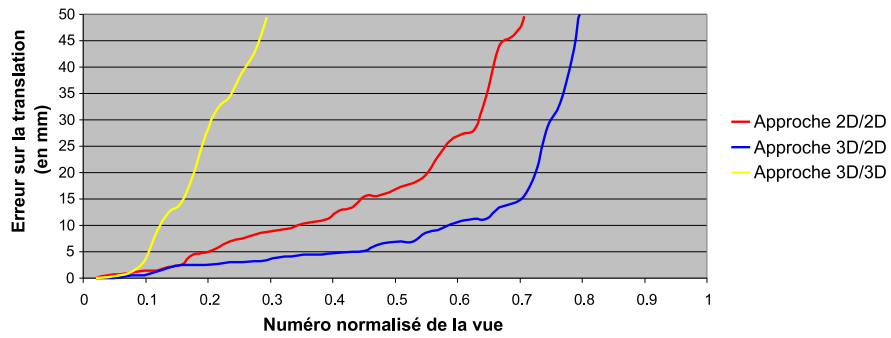
Les figures 6.13 et 6.14 présentent respectivement les erreurs de position et d'orientation commises lors de l'estimation de la pose des différents objets.

D'une manière générale, on peut constater que les solutions 2D/2D et 3D/2D déterminent, dans une large majorité des cas, la pose de l'objet avec une erreur sur la position inférieure au centimètre et une erreur sur l'orientation inférieure au degré. Pour la solution 3D/3D, les erreurs commises sont généralement plus importantes, avec des erreurs sur la position de plusieurs centimètres, voir dizaines de centimètres, et des erreurs sur l'orientation de plusieurs degrés.

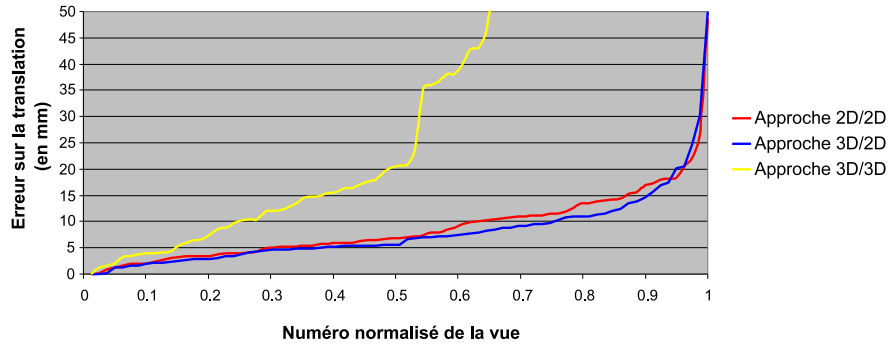
La précision des poses obtenues avec les méthodes 2D/2D et 3D/2D sont relativement similaires. Dans 75% des vues de la culasse et de la portière et 60% des vues de l'étalon, l'erreur commise sur la position est inférieure au centimètre alors que l'erreur commise sur l'orientation est inférieure à un degré. Si l'on compare l'erreur de position à la distance d'observation, on constate que l'erreur commise sur la position est inférieure à 5% de la distance d'observation pour la totalité des vues de la culasse et de la portière, et pour environ 75% des vues de l'étalon.



(a) Résultats obtenus avec la portière

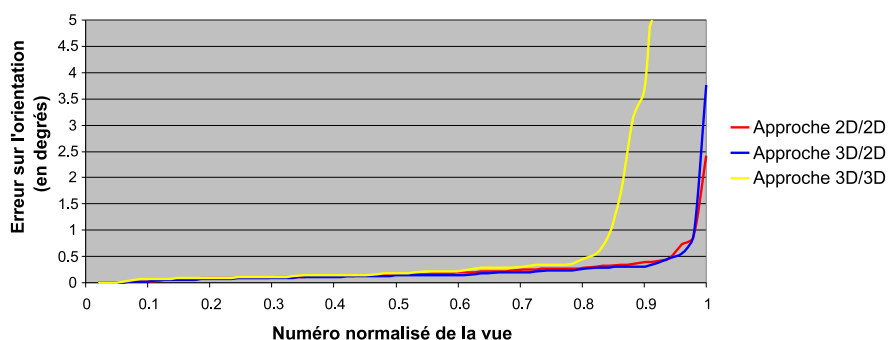


(b) Résultats obtenus avec l'étalon

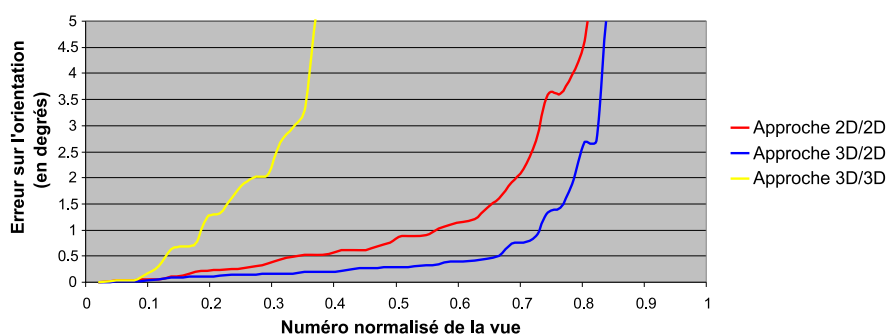


(c) Résultats obtenus avec la culasse

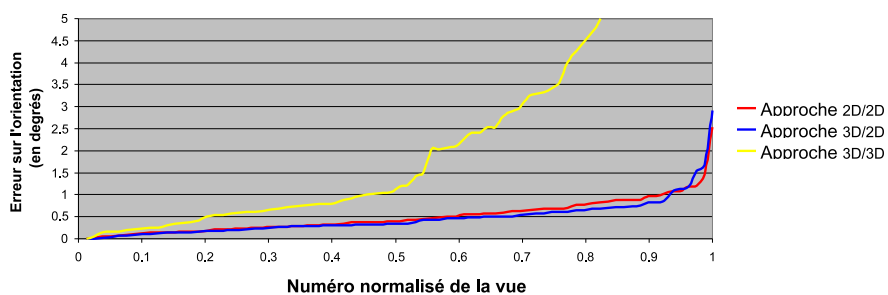
FIG. 6.13 – Erreurs de translation pour les différentes vues de la portière, de l'étalon et de la culasse. Le graphe (a) représente l'erreur commise sur la translation pour chacune des 50 vues de la portière, celles-ci étant triées par ordre croissant d'erreur. Les vues sont indicées selon une numérotation normalisée : la vue ayant l'erreur minimale correspond au numéro 0, la vue ayant l'erreur maximale correspond au numéro 1 et la vue ayant l'erreur médiane correspond à l'abscisse 0.5. De la même manière, les graphes (b) et (c) représentent respectivement l'erreur commise sur la translation dans les différentes vues de l'étalon et de la culasse.



(a) Résultats obtenus avec la portière



(b) Résultats obtenus avec l'étalon



(c) Résultats obtenus avec la culasse

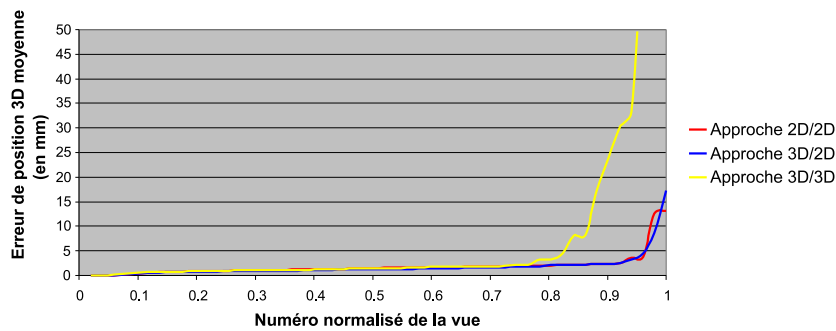
FIG. 6.14 – Erreurs de rotation. *Le graphe (a) représente l'erreur commise sur la rotation pour chacune des 50 vues de la portière, celles-ci étant triées par ordre croissant d'erreur. Les vues sont indicées selon une numérotation normalisée : la vue ayant l'erreur minimale correspond au numéro 0, la vue ayant l'erreur maximale correspond au numéro 1 et la vue ayant l'erreur médiane correspond à l'abscisse 0.5. De la même manière, les graphes (b) et (c) représentent respectivement l'erreur commise sur la rotation dans les différentes vues de l'étalon et de la culasse.*

6.3.4 Précision de la position 3D des points de surface visibles

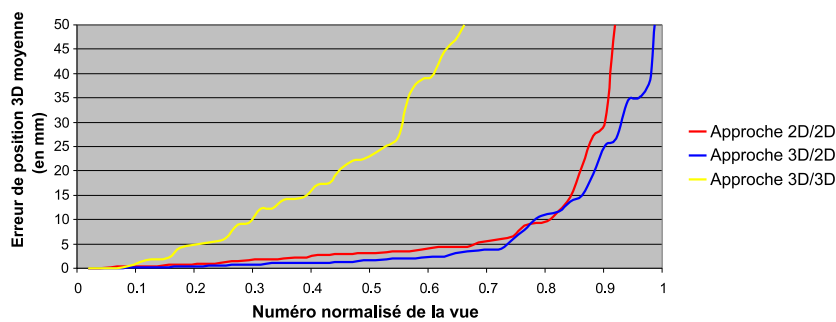
Les résultats obtenus avec la portière, l'étalon et la culasse sont illustrés dans la figure 6.15.

Concernant les approches 2D/2D et 3D/2D, l'erreur moyenne commise sur la position 3D des points de surface est inférieure à 5mm pour 75% à 95% des vues en fonction de la scène observée. La répartition de cette erreur à la surface de l'objet est généralement relativement uniforme, comme l'illustrent les séries d'images des figures 6.17 à 6.19 pour la portière. Pour chacune des trois méthodes, les images-résultats, correspondant au meilleur recalage de la portière ainsi qu'au pire recalage et à un recalage moyen de celle-ci, sont présentées dans les figures 6.21, 6.23 et 6.25. Les images-résultats du recalage moyen pour les différentes méthodes sont présentées en pleines pages dans les figures 6.28, 6.31 et 6.34. Les résultats obtenus avec la culasse et l'étalon étant relativement similaires, ceux-ci sont présentés dans l'annexe C.

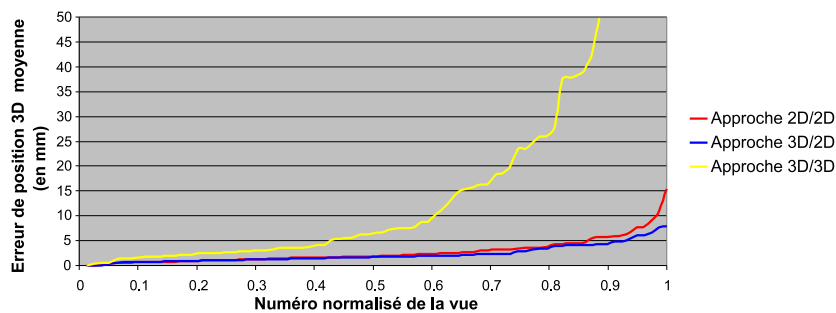
Enfin, on notera que certaines vues de l'étalon présentent une erreur de position 3D moyenne de l'ordre du centimètre alors que l'erreur de reprojection moyenne associée est de l'ordre 1 à 2 pixels. La figure 6.16 correspond à une vue pour laquelle l'erreur de reprojection varie entre 0.5 et 2.5 pixels, avec une erreur moyenne de 1.2 pixels, alors que l'erreur de position 3D des points de surface varie entre 2.5cm et 3.3cm, avec une erreur moyenne de 3cm. On peut voir dans la figure 6.16 qu'il s'agit d'un cas de figure où l'erreur 3D correspond principalement à une translation le long de la droite passant par le centre optique de la caméra et le centre de gravité de l'étalon. Au niveau des points de surfaces de l'objet, cette erreur correspond principalement à une translation le long de leur rayon de projection et n'entraîne, par conséquent, qu'une faible erreur de reprojection.



(a) Résultats obtenus avec la portière



(b) Résultats obtenus avec l'étalon



(c) Résultats obtenus avec la culasse

FIG. 6.15 – Erreurs de position 3D des points de surface visibles pour les différentes vues de la portière, de l'étalon et de la culasse. Le graphe (a) représente l'erreur de position 3D des points de surface visibles pour chacune des 50 vues de la portière, celles-ci étant triées par ordre croissant d'erreur. Les vues sont indicées selon une numérotation normalisée : la vue ayant l'erreur minimale correspond au numéro 0, la vue ayant l'erreur maximale correspond au numéro 1 et la vue ayant l'erreur médiane correspond à l'abscisse 0.5. De la même manière, les graphes (b) et (c) représentent respectivement les erreurs de position 3D des points de surface de l'étalon et de la culasse

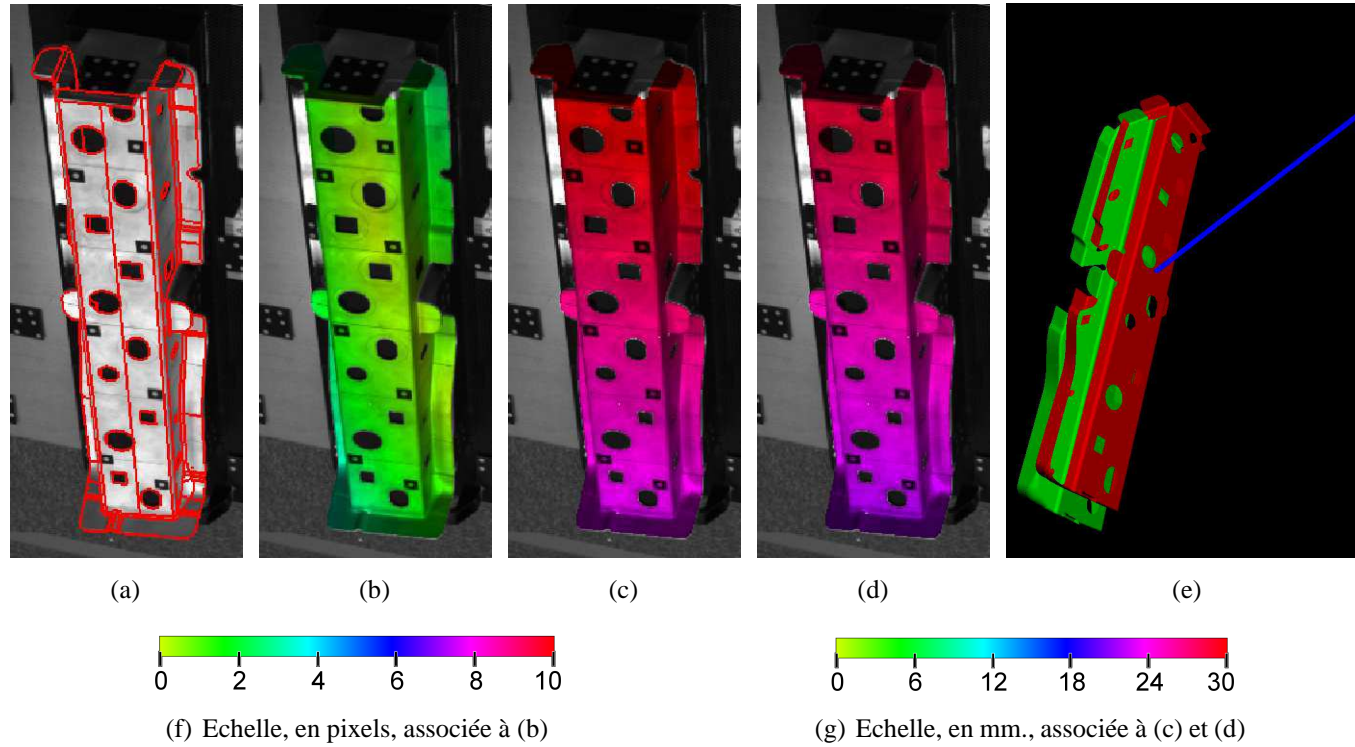


FIG. 6.16 – Cas d’une erreur de reprojection faible malgré une erreur 3D non négligeable. (a) : *reprojection du modèle CAO*. (b) : *l’erreur de reprojection (échelle = fig.(f))*. (c) : *l’erreur de position 3D des points de surface (échelle = fig.(g))*. (d) : *l’erreur de position 3D des points de surface le long des rayon de projection uniquement (échelle = fig.(g))*. (e) : *vue 3D de la scène ; en vert, la pose estimée ; en rouge la pose réelle ; en bleu, la droite liant le centre de gravité de l’étalon rouge au centre optique de la caméra*. On observe que l’erreur de reprojection est de l’ordre de 1 pixel alors que l’erreur de position 3D varie de 2 à 3 cm. En comparant (c) et (d), on constate que l’erreur de position 3D correspond principalement à une erreur le long des rayons de projection.

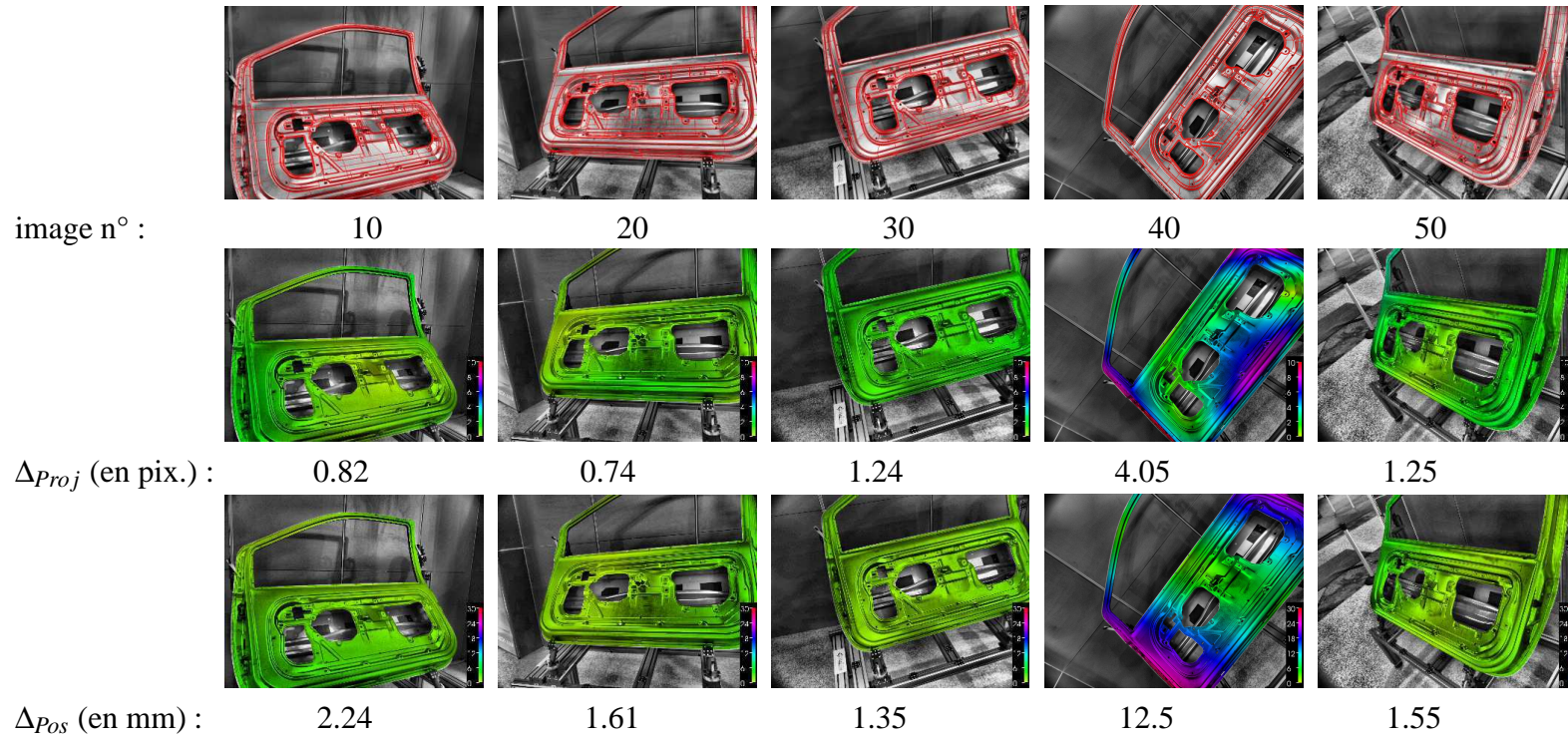


FIG. 6.17 – Exemples de résultats obtenus sur la portière avec la méthode 2D/2D. La première ligne d'images représente la reprojection de la portière (en rouge). La seconde ligne représente la distribution de l'erreur de reprojection (le nombre indiqué sous chaque vue correspond à l'erreur de reprojection moyenne en pixels). La dernière ligne représente la distribution de l'erreur de position 3D des points de surface (le nombre indiqué sous chaque vue correspond à l'erreur 3D moyenne en millimètres).

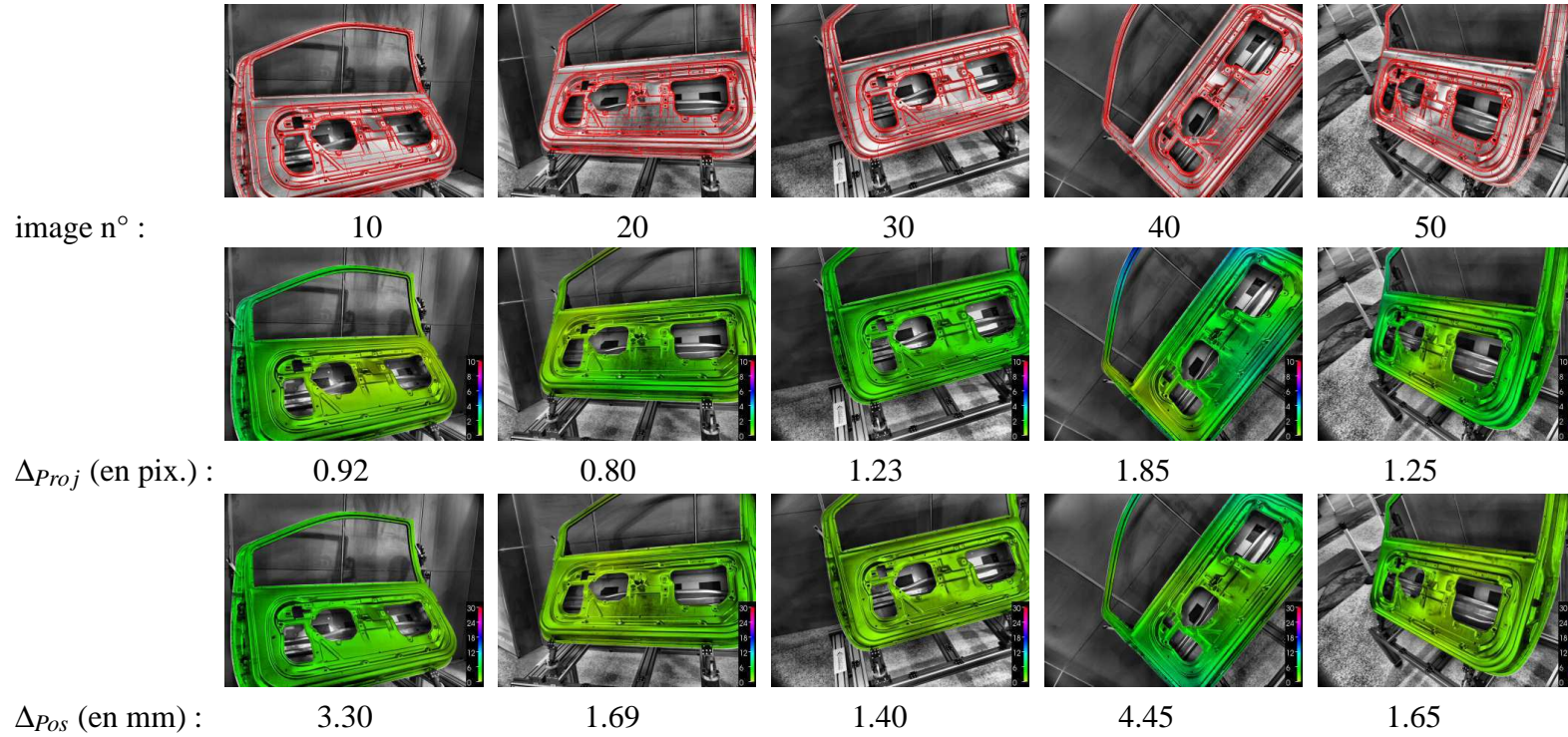


FIG. 6.18 – Exemples de résultats obtenus sur la portière avec la méthode 3D/2D. La première ligne d'images représente la reprojection de la portière (en rouge). La seconde ligne représente la distribution de l'erreur de reprojection (le nombre indiqué sous chaque vue correspond à l'erreur de reprojection moyenne en pixels). La dernière ligne représente la distribution de l'erreur de position 3D des points de surface (le nombre indiqué sous chaque vue correspond à l'erreur 3D moyenne en millimètres).

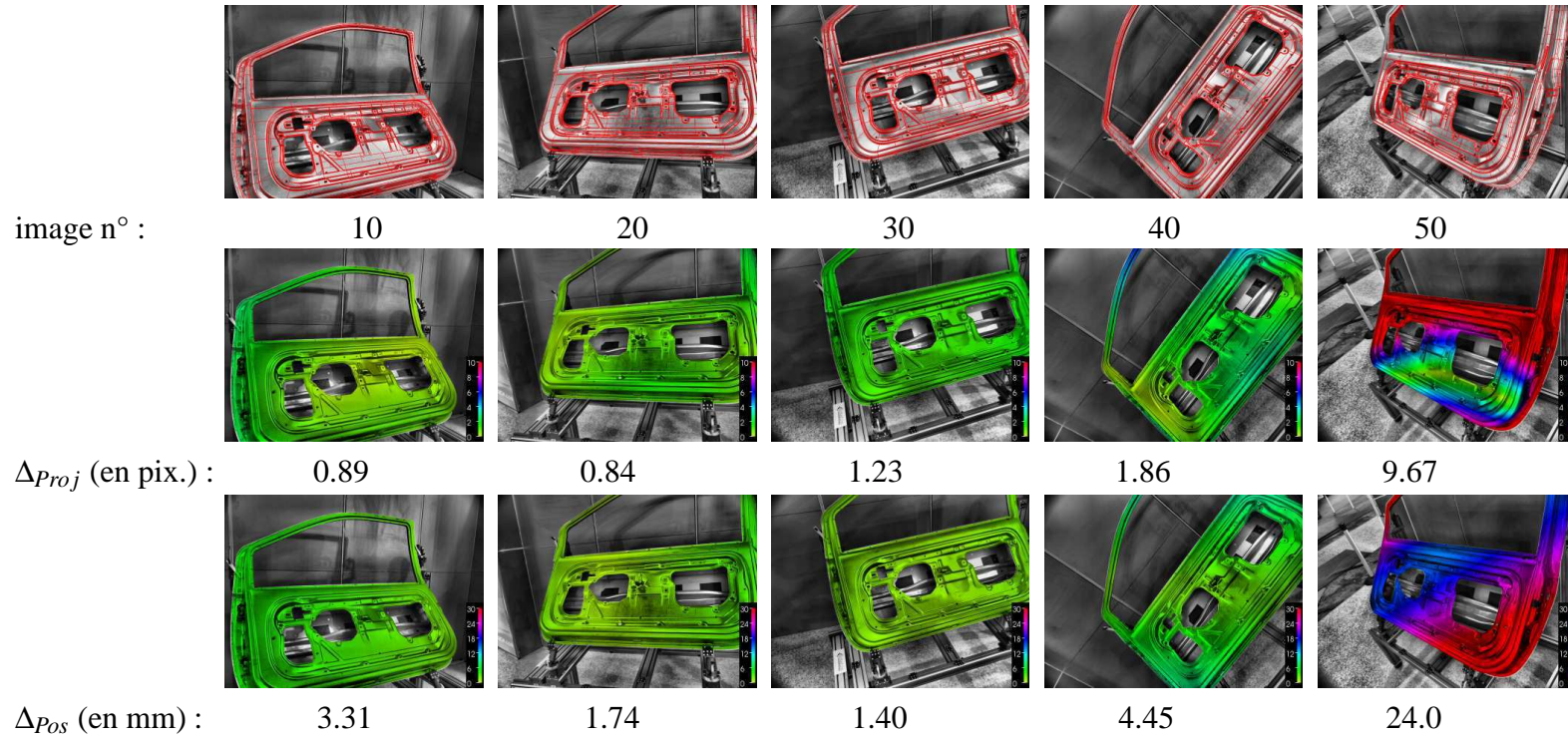
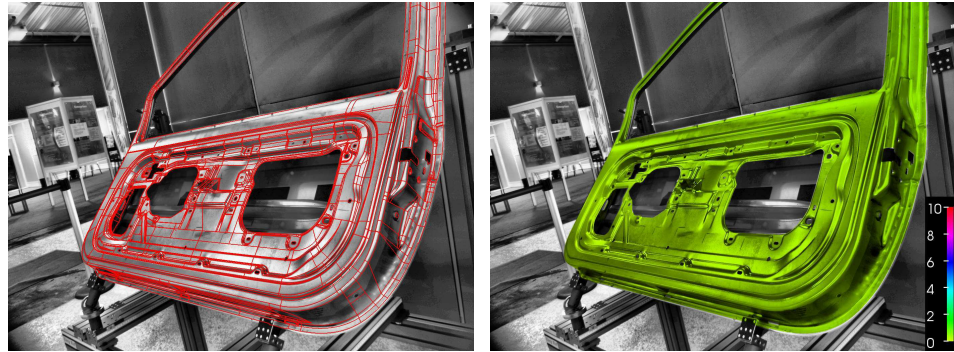


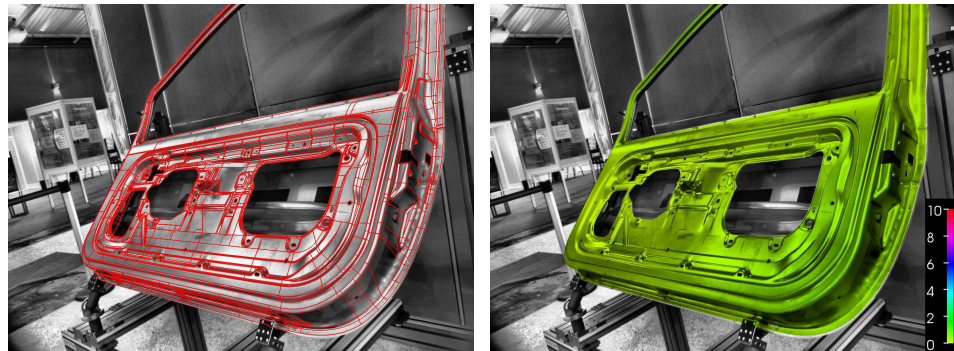
FIG. 6.19 – Exemples de résultats obtenus sur la portière avec la méthode 3D/3D. La première ligne d'images représente la reprojection de la portière (en rouge). La seconde ligne représente la distribution de l'erreur de reprojection (le nombre indiqué sous chaque vue correspond à l'erreur de reprojection moyenne en pixels). La dernière ligne représente la distribution de l'erreur de position 3D des points de surface (le nombre indiqué sous chaque vue correspond à l'erreur 3D moyenne en millimètres).

Reprojection du modèle

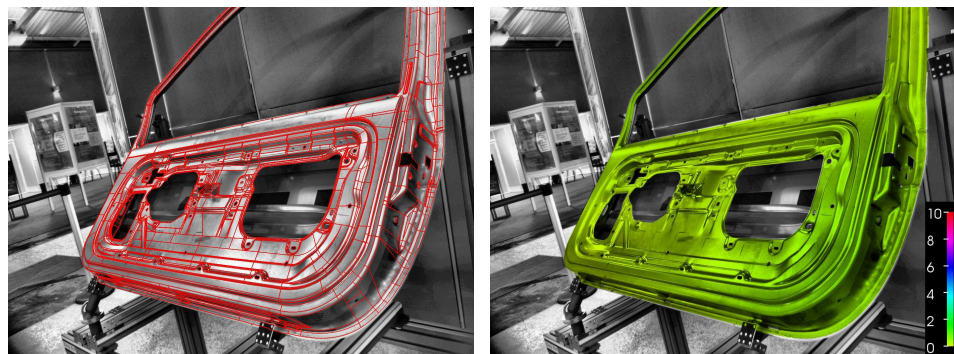
Erreur de reprojection



Méthode 2D/2D



Méthode 3D/2D

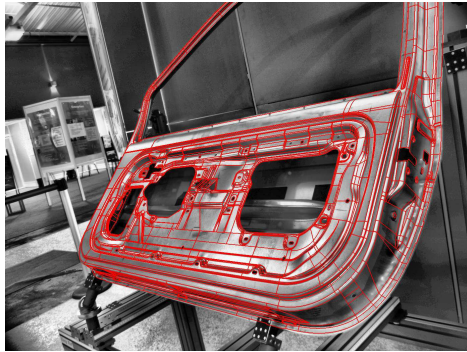


Méthode 3D/3D

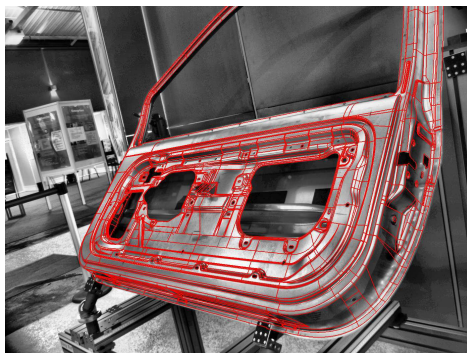
FIG. 6.20 – Erreur de reprojection du meilleur recalage de la portière. *Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente, en fausses couleurs, l'erreur de reprojection associée (en pixels).*

Reprojection du modèle

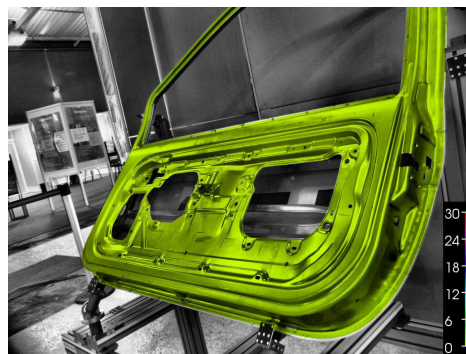
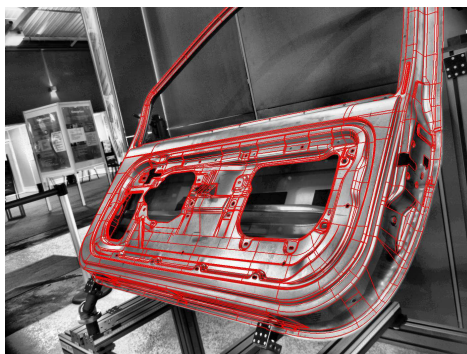
Erreur de position 3D



Méthode 2D/2D



Méthode 3D/2D

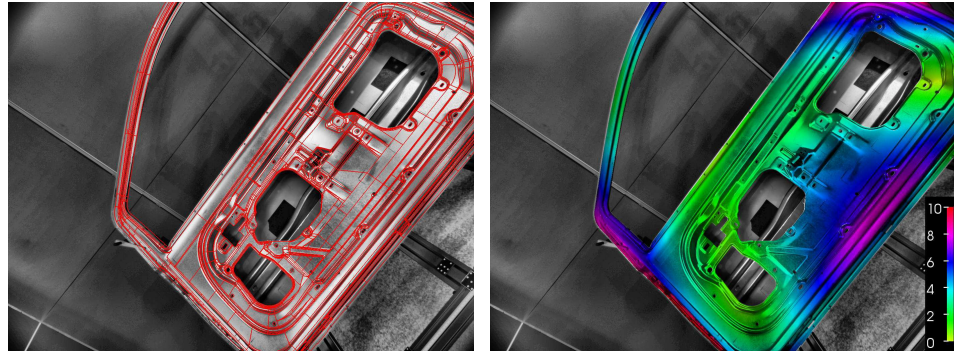


Méthode 3D/3D

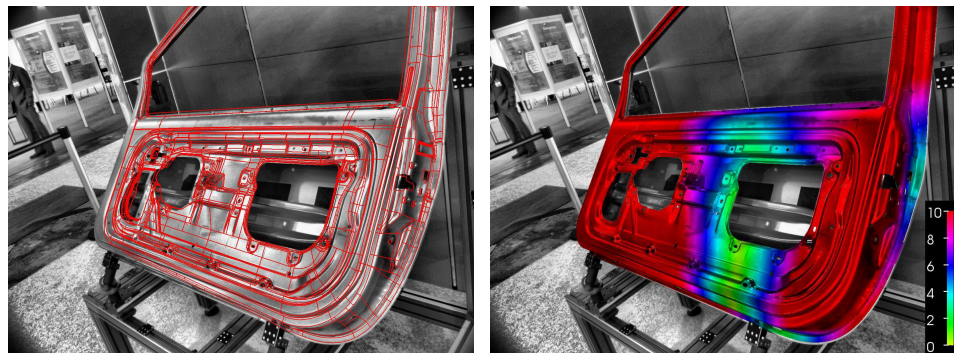
FIG. 6.21 – Erreur de position 3D des points de surface du meilleur recalage de la portière. Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de position 3D des points de surface (en mm).

Reprojection du modèle

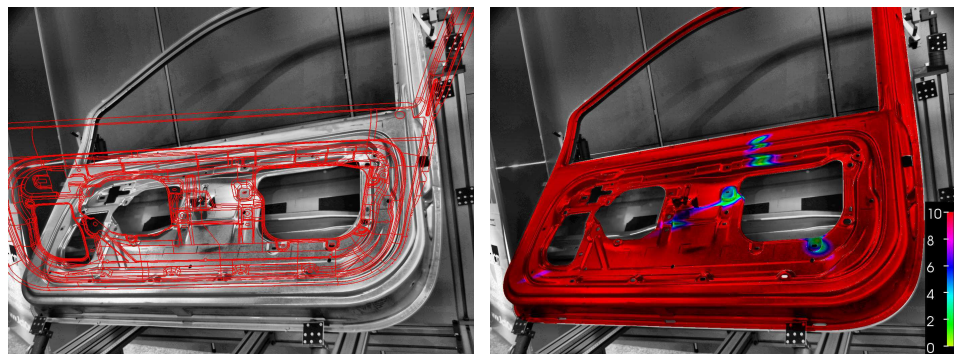
Erreur de reprojection



Méthode 2D/2D



Méthode 3D/2D

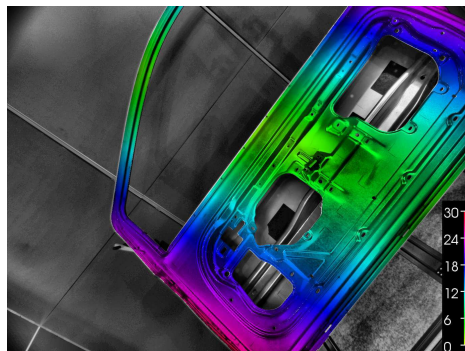


Méthode 3D/3D

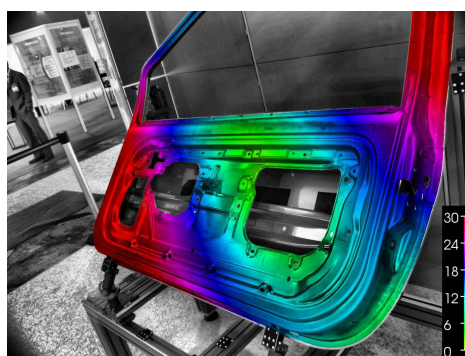
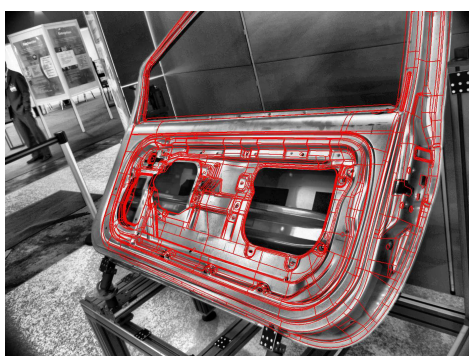
FIG. 6.22 – Erreur de reprojection du plus mauvais recalage de la portière. *Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de reprojection associée (en pixels).*

Reprojection du modèle

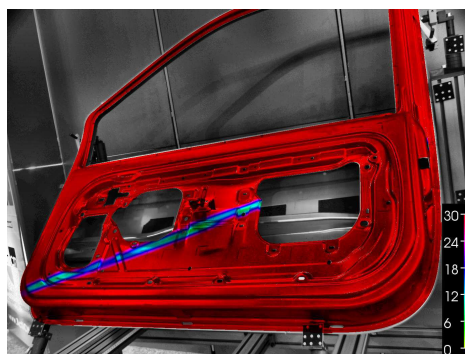
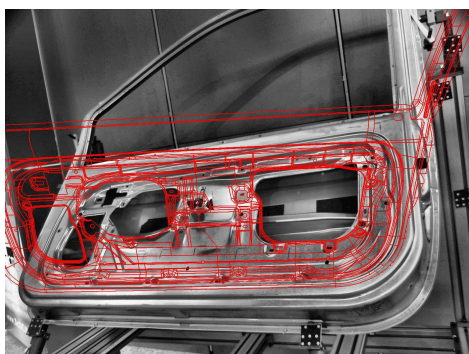
Erreur de position 3D



Méthode 2D/2D



Méthode 3D/2D

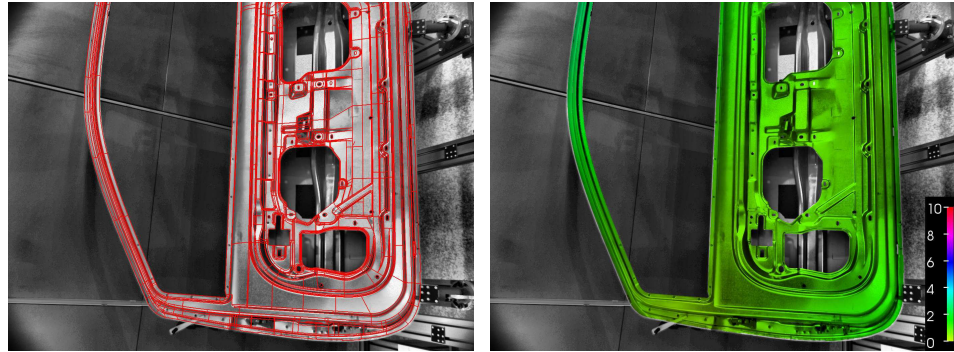


Méthode 3D/3D

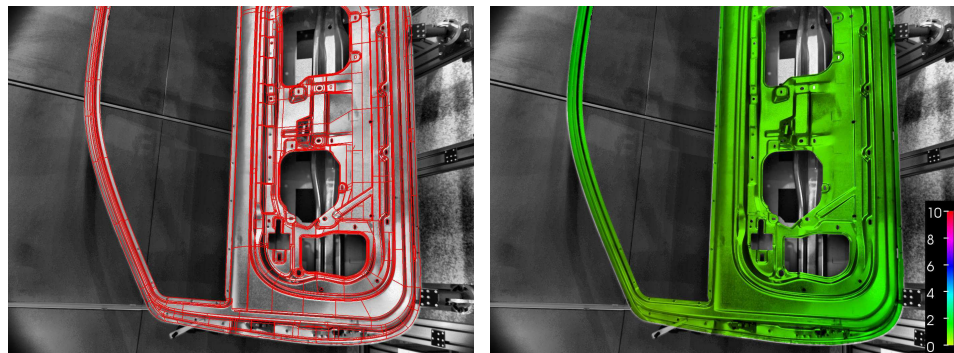
FIG. 6.23 – Erreur de position 3D des points de surface du plus mauvais recalage de la portière. Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de position 3D des points de surface (en mm).

Reprojection du modèle

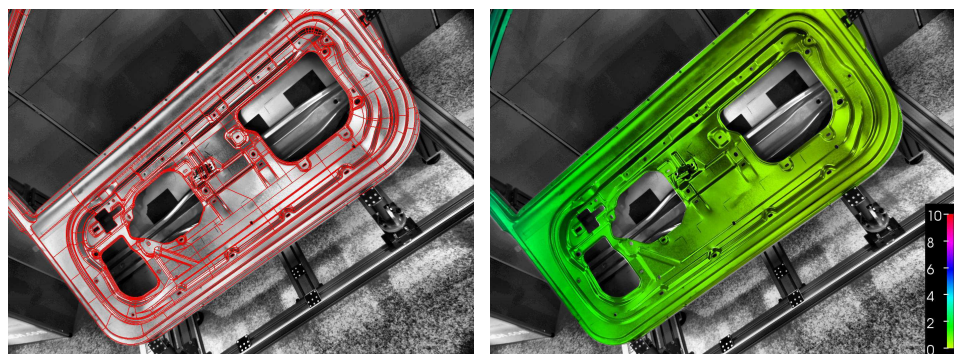
Erreur de reprojection



Méthode 2D/2D



Méthode 3D/2D

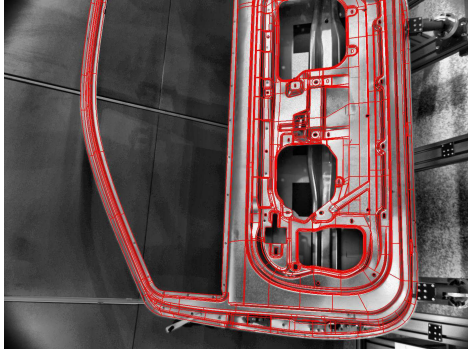


Méthode 3D/3D

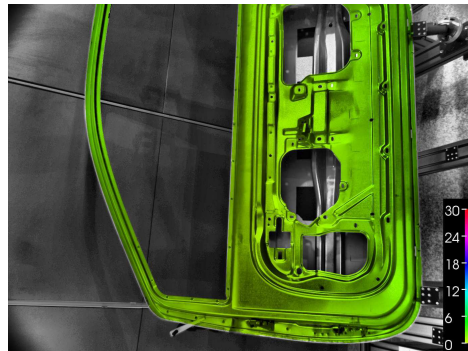
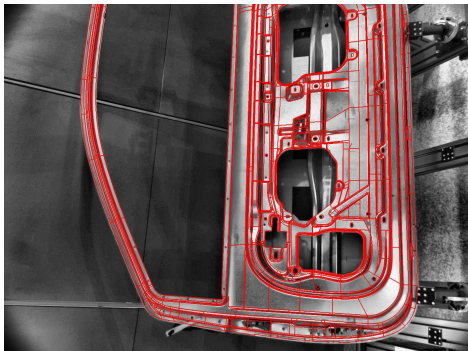
FIG. 6.24 – Erreur de reprojection pour un recalage moyen de la portière. *Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de reprojection associée (en pixels).*

Reprojection du modèle

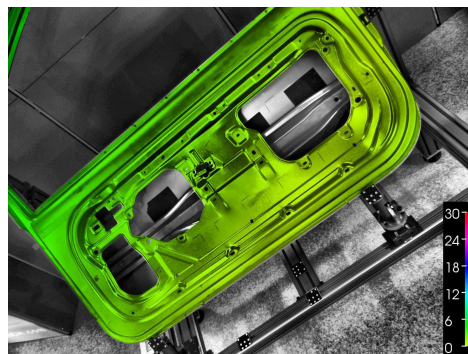
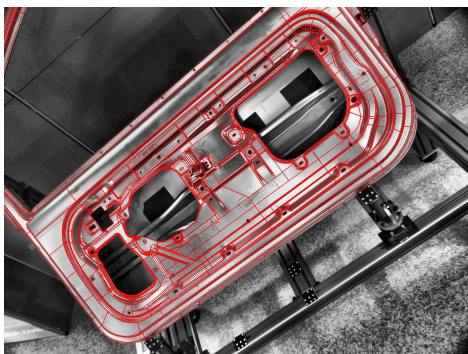
Erreur de position 3D



Méthode 2D/2D



Méthode 3D/2D



Méthode 3D/3D

FIG. 6.25 – Erreur de position 3D des points de surface pour un recalage moyen de la portière. *Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de position 3D des points de surface (en mm).*

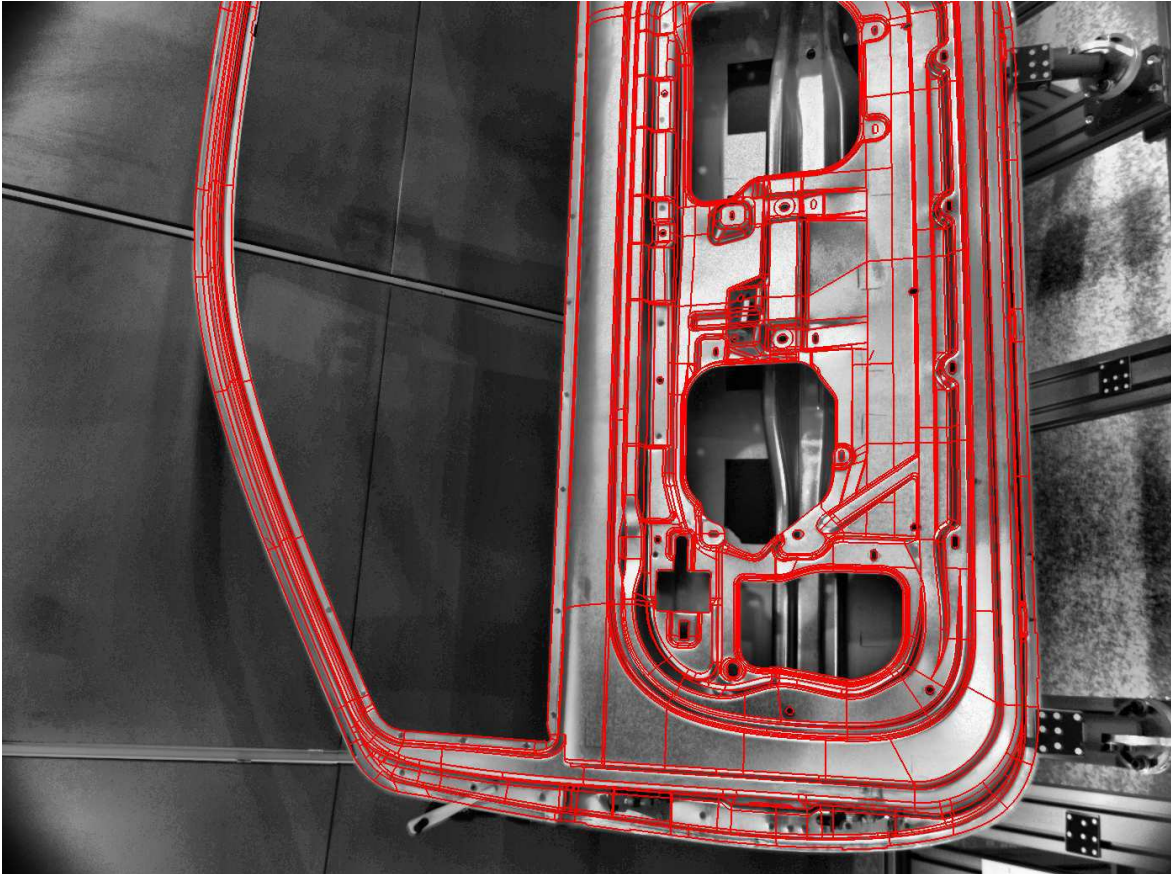


FIG. 6.26 – Reprojection du modèle de la portière pour un recalage moyen selon la méthode 2D/2D.

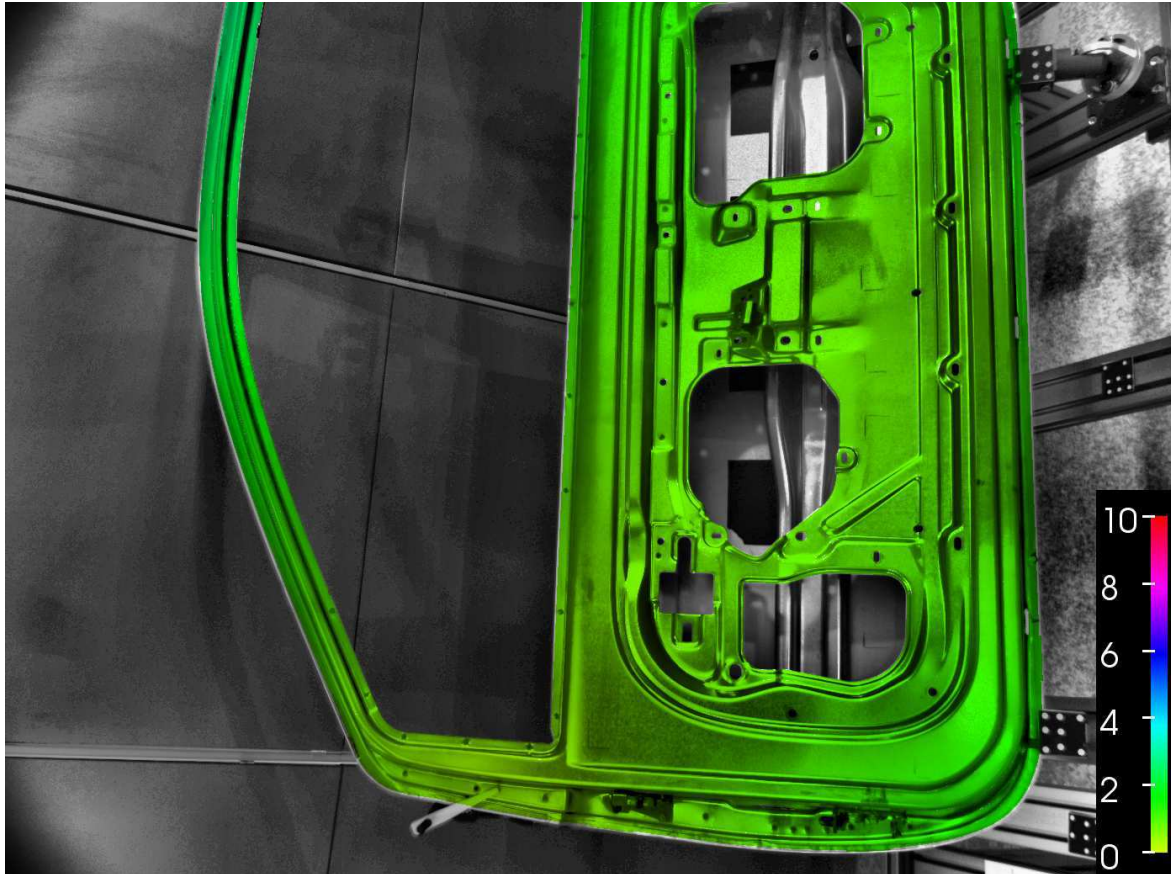


FIG. 6.27 – Erreur de reprojection à la surface de la portière pour un recalage moyen selon la méthode 2D/2D. *L'échelle est donnée en pixels.*

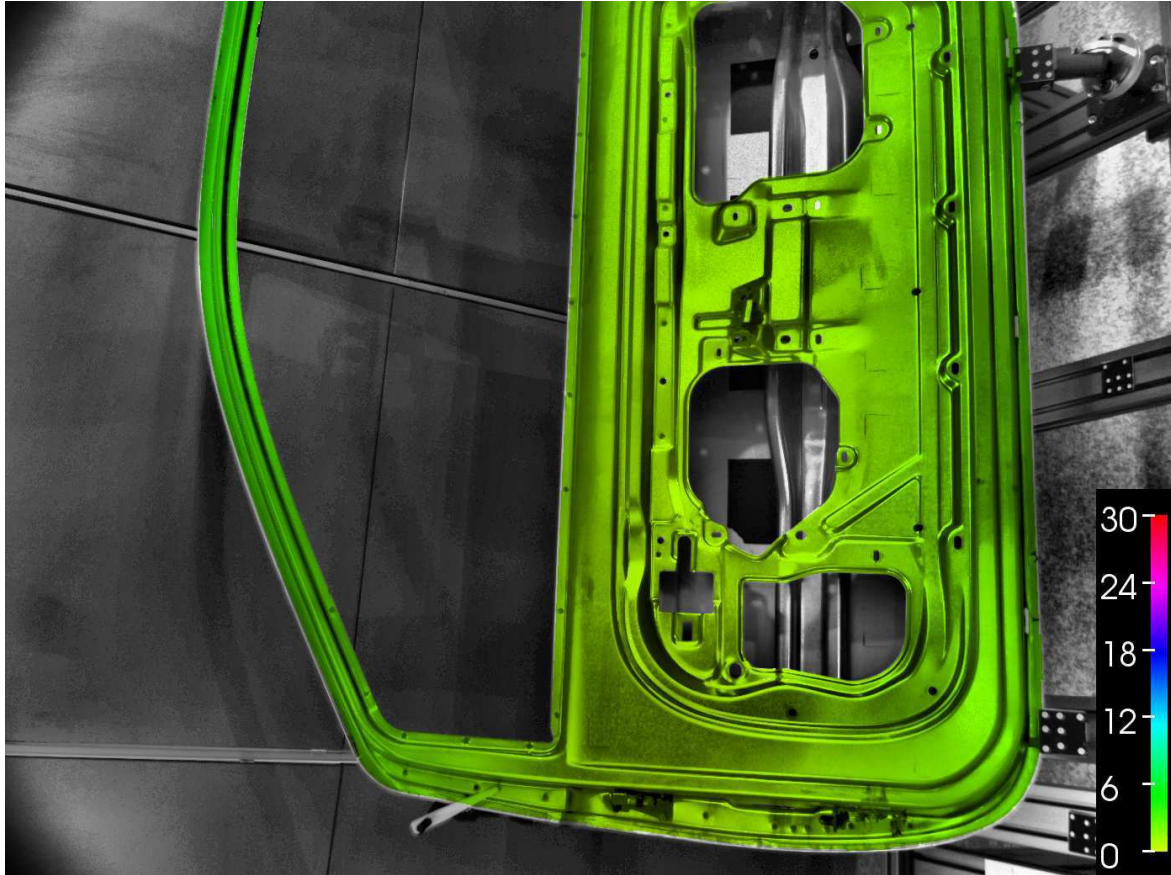


FIG. 6.28 – Erreur de position des points de surface de la portière pour un recalage moyen selon la méthode 2D/2D. *L'échelle est donnée en mm.*

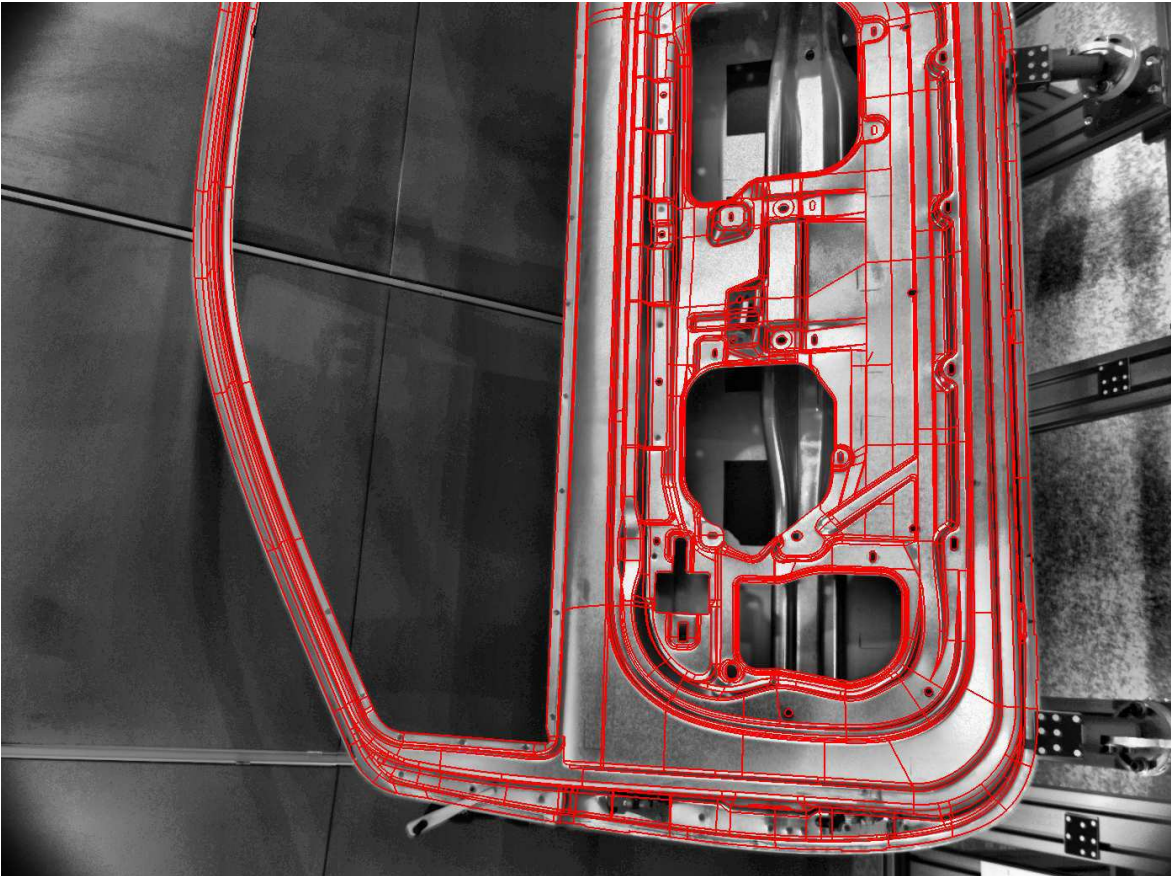


FIG. 6.29 – Reprojection du modèle de la portière pour un recalage moyen selon la méthode 3D/2D.

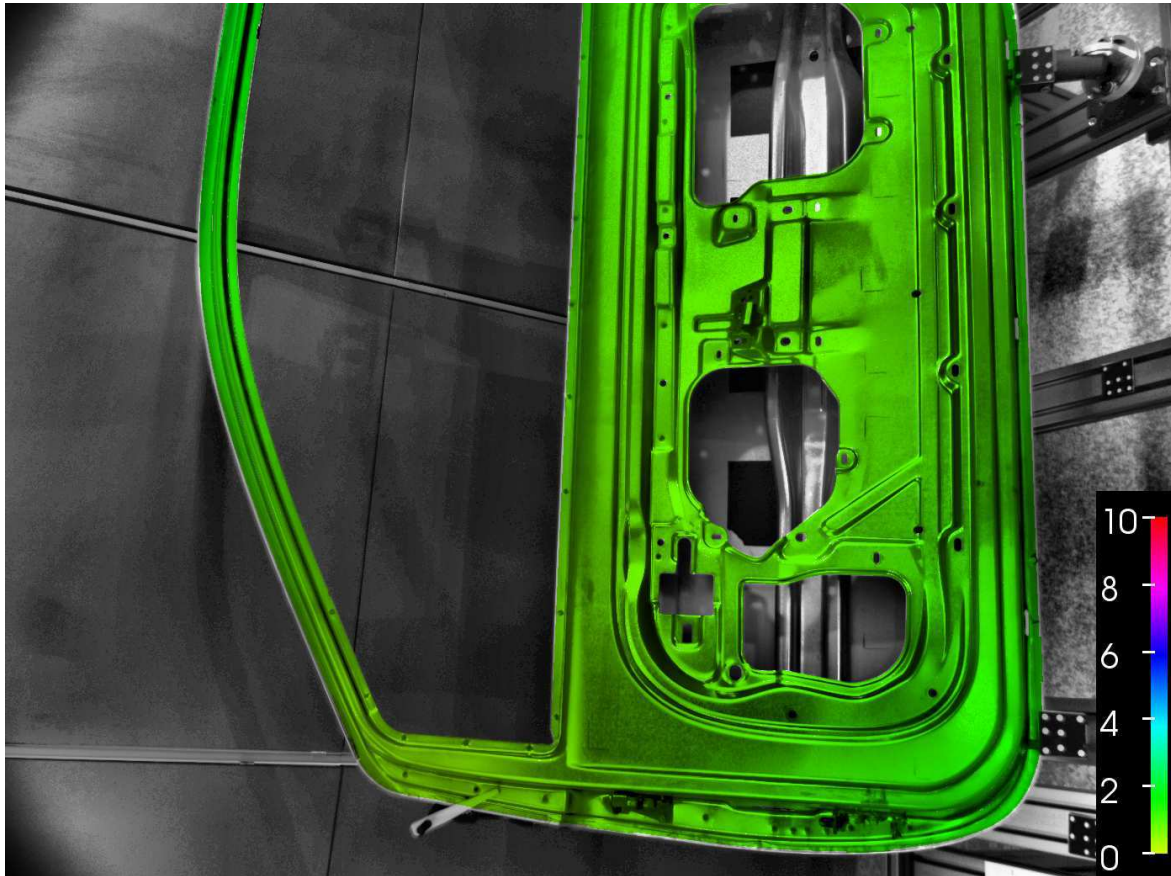


FIG. 6.30 – Erreur de reprojection à la surface de la portière pour un recalage moyen selon la méthode 3D/2D. *L'échelle est donnée en pixels.*



FIG. 6.31 – Erreur de position des points de surface de la portière pour un recalage moyen selon la méthode 3D/2D. *L'échelle est donnée en mm.*

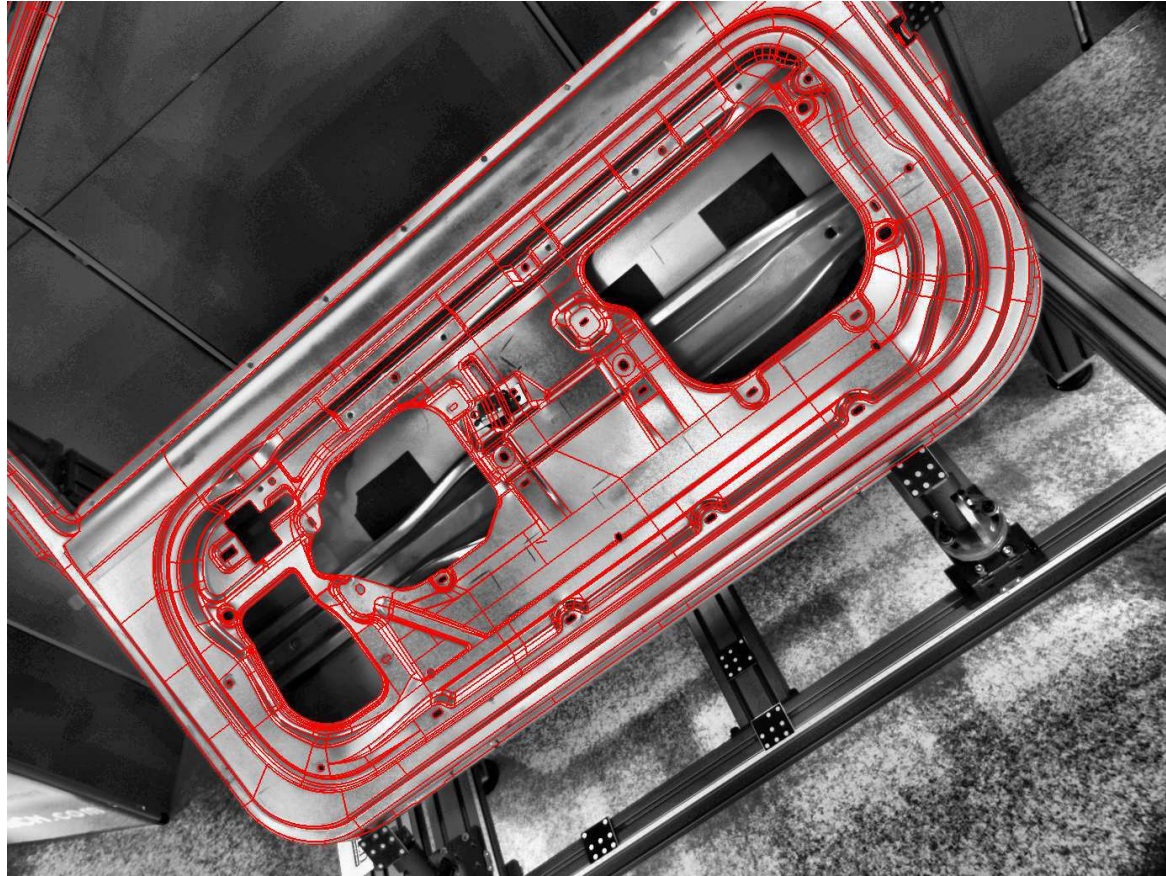


FIG. 6.32 – Reprojection du modèle de la portière pour un recalage moyen selon la méthode 3D/3D.

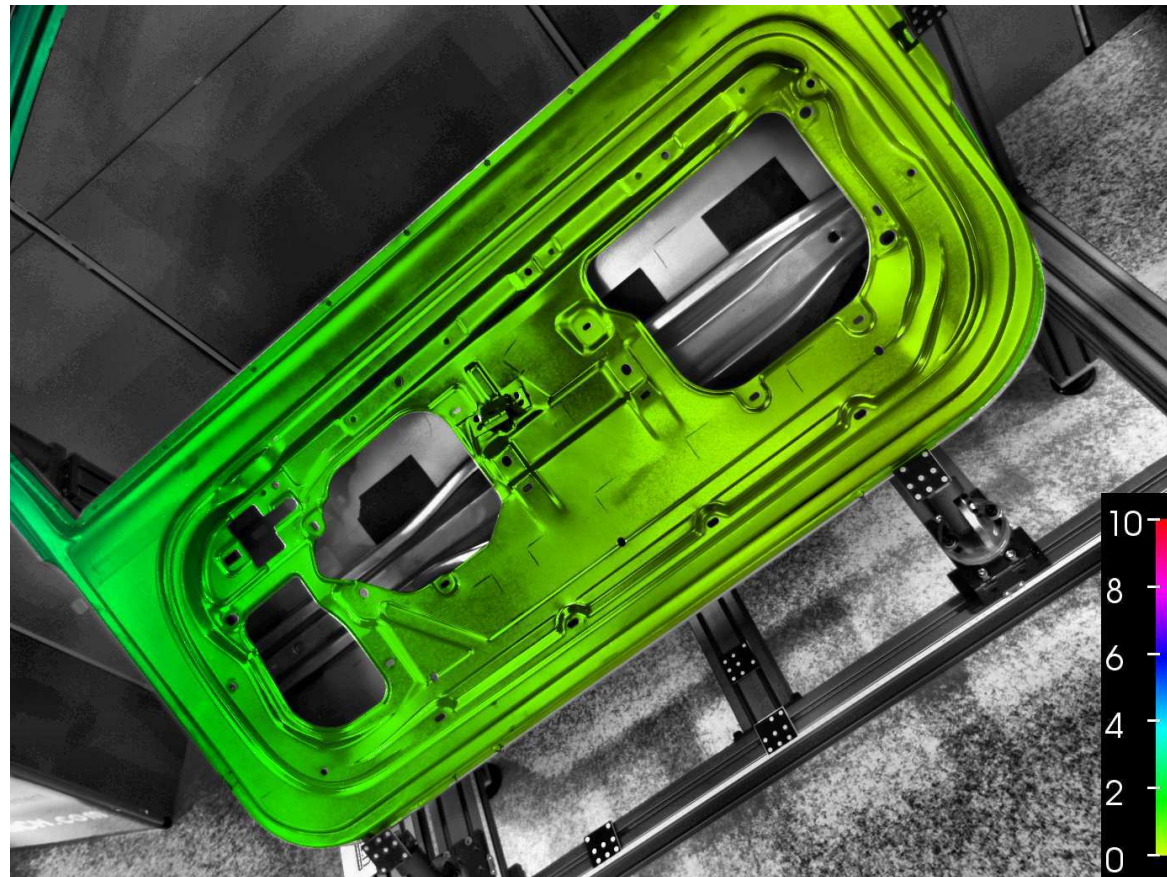


FIG. 6.33 – Erreur de reprojection à la surface de la portière pour un recalage moyen selon la méthode 3D/3D. *L'échelle est donnée en pixels.*

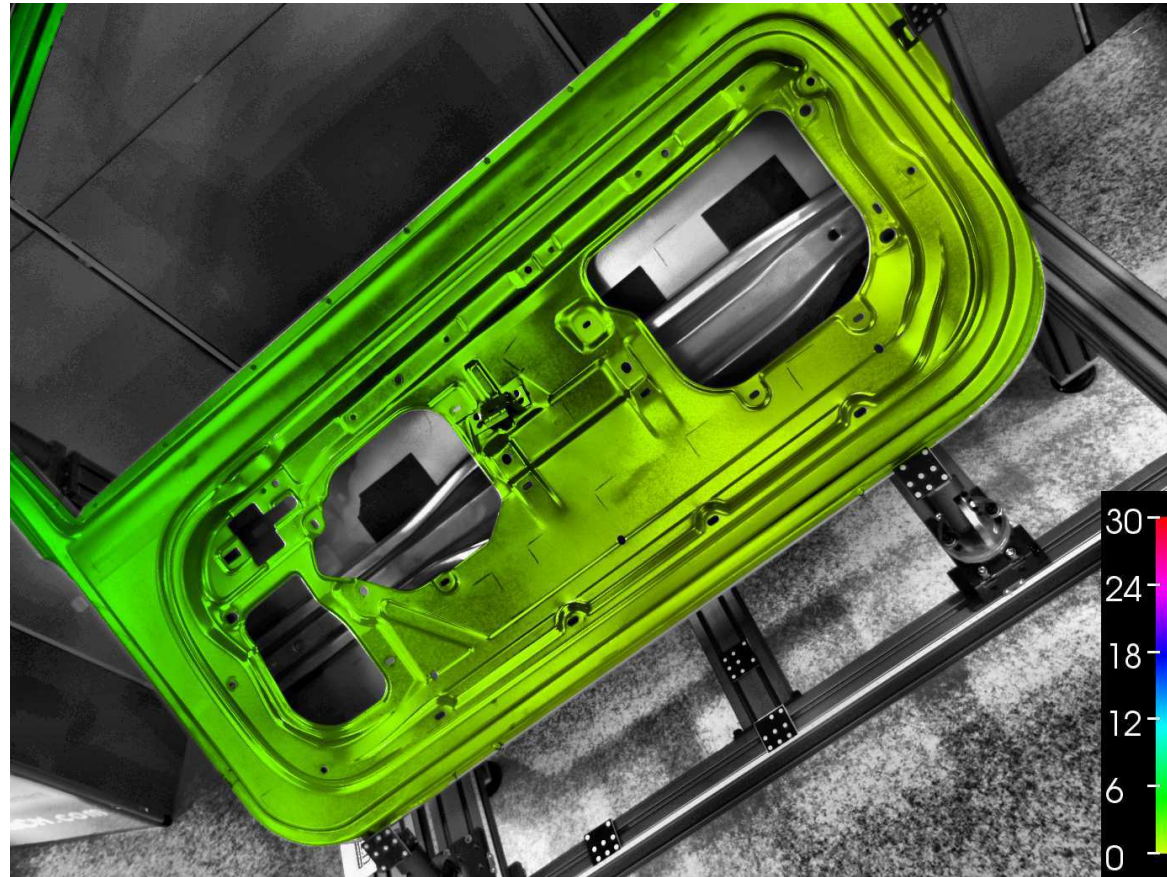


FIG. 6.34 – Erreur de position des points de surface de la portière pour un recalage moyen selon la méthode 3D/3D. L'échelle est donnée en mm.

6.4 Tests dans un environnement non contrôlé

Cette expérience a pour but de tester l'algorithme dans le cadre d'un environnement non contrôlé. Les images traitées ont donc été acquises dans des conditions totalement différentes de celles utilisées pour élaborer le modèle, que ce soit :

- au niveau du capteur : lors de l'expérience, le système d'acquisition vidéo utilisé est différent (résolution différente, focale différente) et calibré très approximativement ;
- au niveau des conditions d'illumination : l'éclairage n'est pas contrôlé ;
- au niveau du point de vue : les images ont été acquises pour des points de vue très différents de ceux utilisés pour créer le modèle (important facteur d'échelle, régions de l'objet n'apparaissant pas dans les images d'apprentissages ...) ;
- au niveau de l'environnement : l'environnement présente des objets parasites comportant de nombreux alésages et l'objet peut être partiellement occulté.

La base d'images utilisée pour cette évaluation est présentée dans la figure 6.35. Le modèle utilisé lors de cette expérience est identique à celui utilisé dans l'expérience précédente, c'est à dire un modèle établi à partir des 3 vues de la figure 6.1. Ne disposant pas de vérité terrain pour ces scènes, la qualité du recalage est laissée à l'appréciation visuelle du lecteur. Les résultats obtenus avec les approches 2D/2D, 3D/2D et 3D/3D sont illustrées dans les figures 6.36, 6.37 et 6.38.

Comme précédemment, les solutions 2D/2D et 3D/2D présentent des résultats similaires. Dans les deux cas, la totalité des vues sont correctement recalées. La perte de précision, comparativement à celle obtenue lors des expériences précédentes, s'explique essentiellement par le calibrage très approximatif de la caméra. Cette expérience souligne la robustesse du recalage dans un environnement non contrôlé.

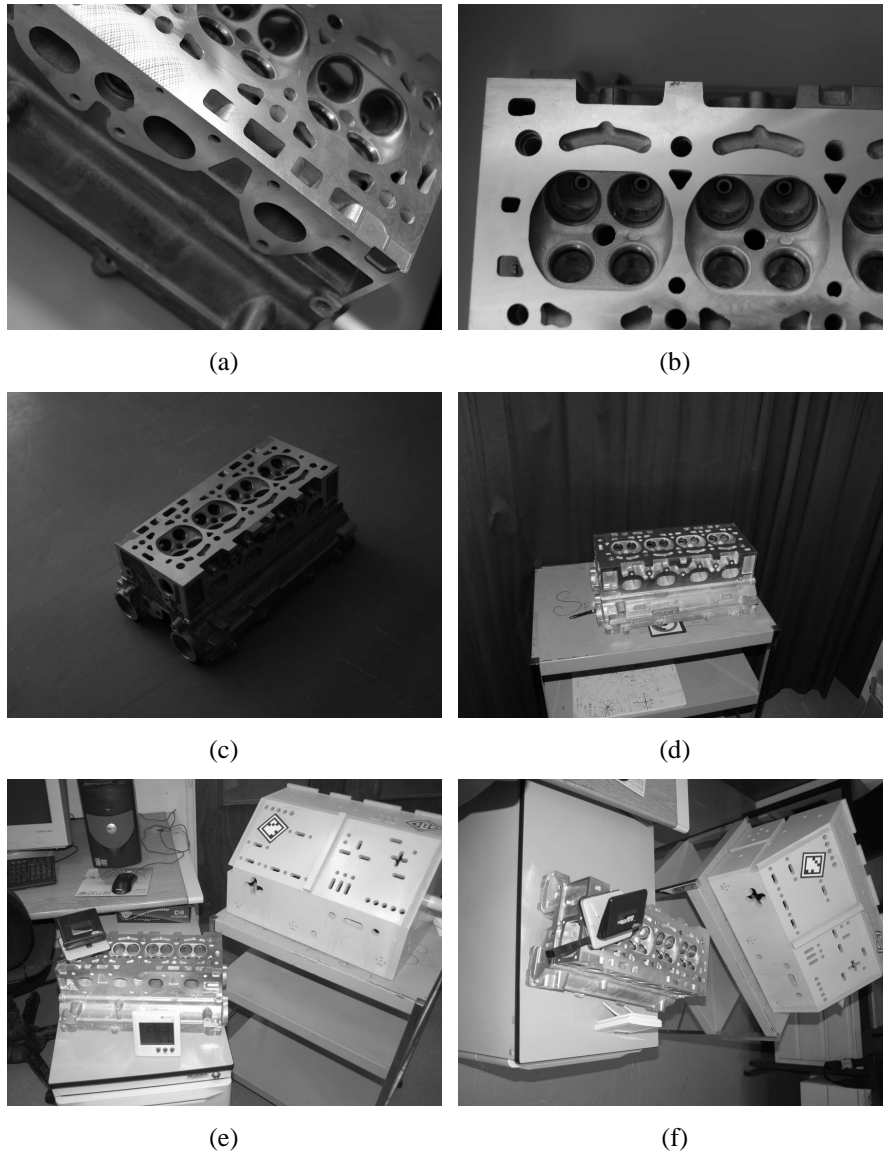


FIG. 6.35 – Base d’images pour l’évaluation en environnement non contrôlé. Dans les images (a) et (b), la culasse n’est que partiellement visible et présente un facteur d’échelle important par rapport aux images d’apprentissages. Les images (c) et (d) correspondent à des changements de point de vues importants (on observe notamment la face arrière qui n’était pas visible dans les vues d’apprentissage). Enfin, les image (e) et (f) présentent la culasse dans un environnement comportant de nombreux objets parasites, certains d’entre eux occultant partiellement la culasse. Comme on peut l’observer, l’ensemble de ces images ont été acquises dans des conditions d’illuminations variées.

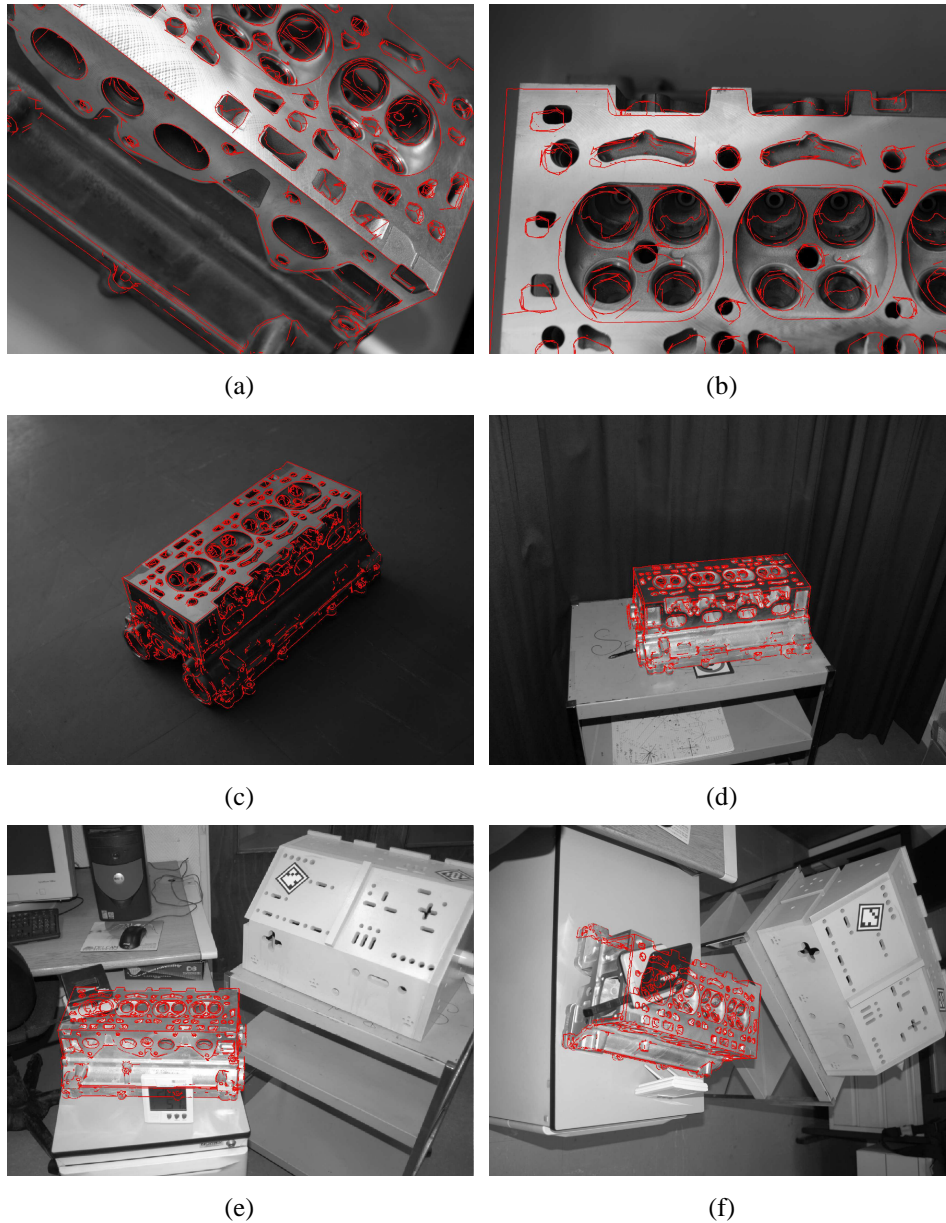


FIG. 6.36 – Recalages selon l'approche 2D/2D dans un environnement non contrôlé. La reprojection du modèle CAO de la culasse est représentée en rouge dans l'image.

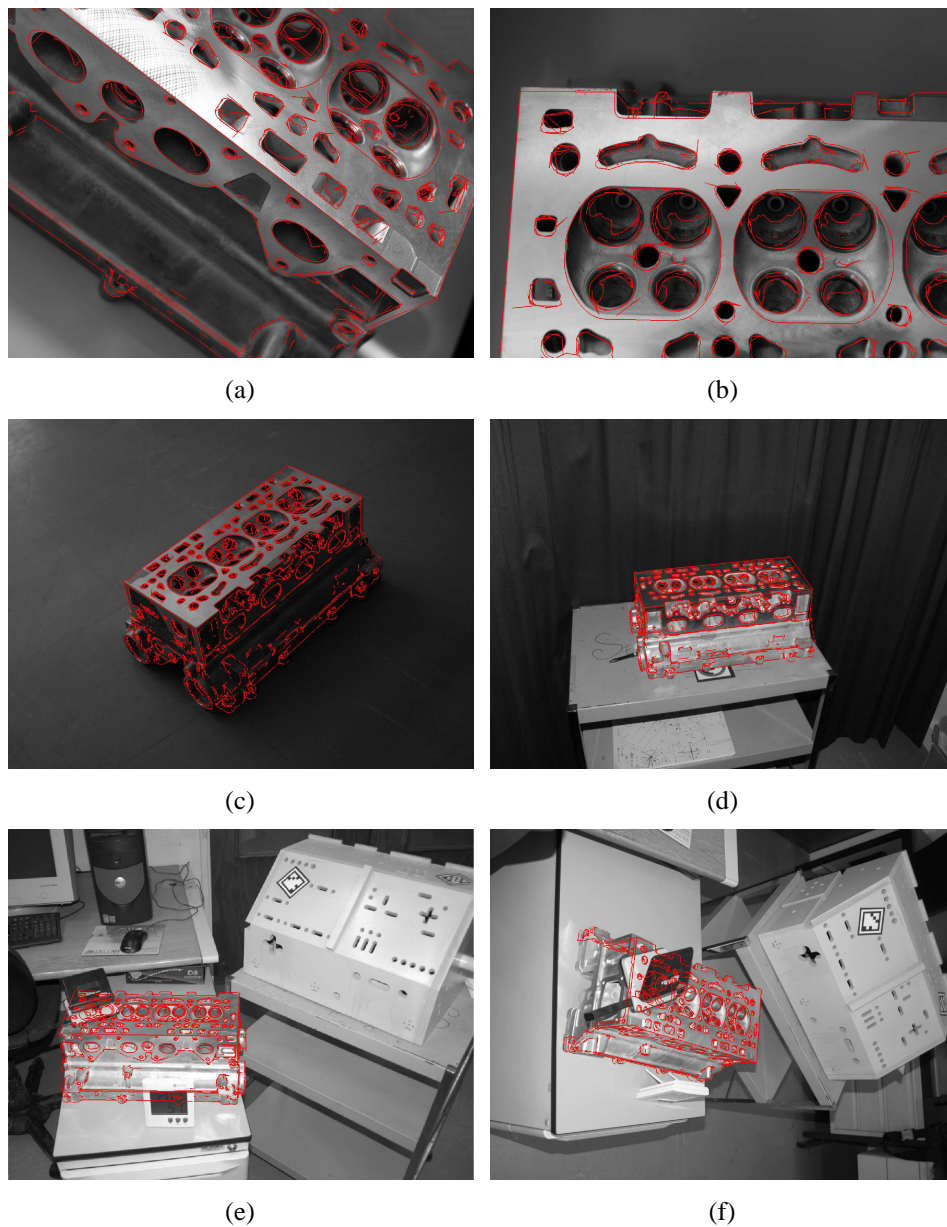


FIG. 6.37 – Recalages selon l'approche 3D/2D dans un environnement non contrôlé. La reprojection du modèle CAO de la culasse est représentée en rouge dans l'image.

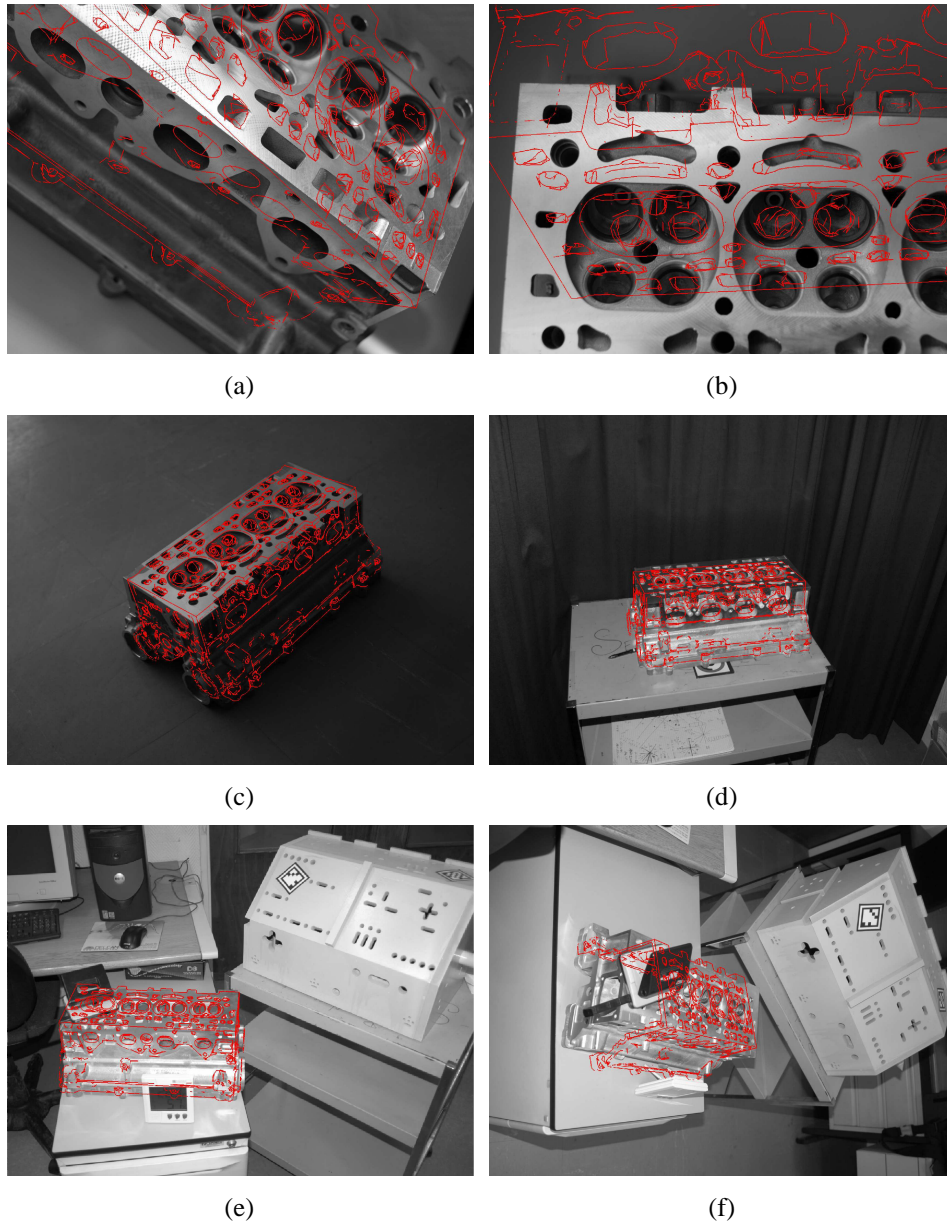


FIG. 6.38 – Recalages selon l'approche 3D/3D dans un environnement non contrôlé. La reprojection du modèle CAO de la culasse est représentée en rouge dans l'image.

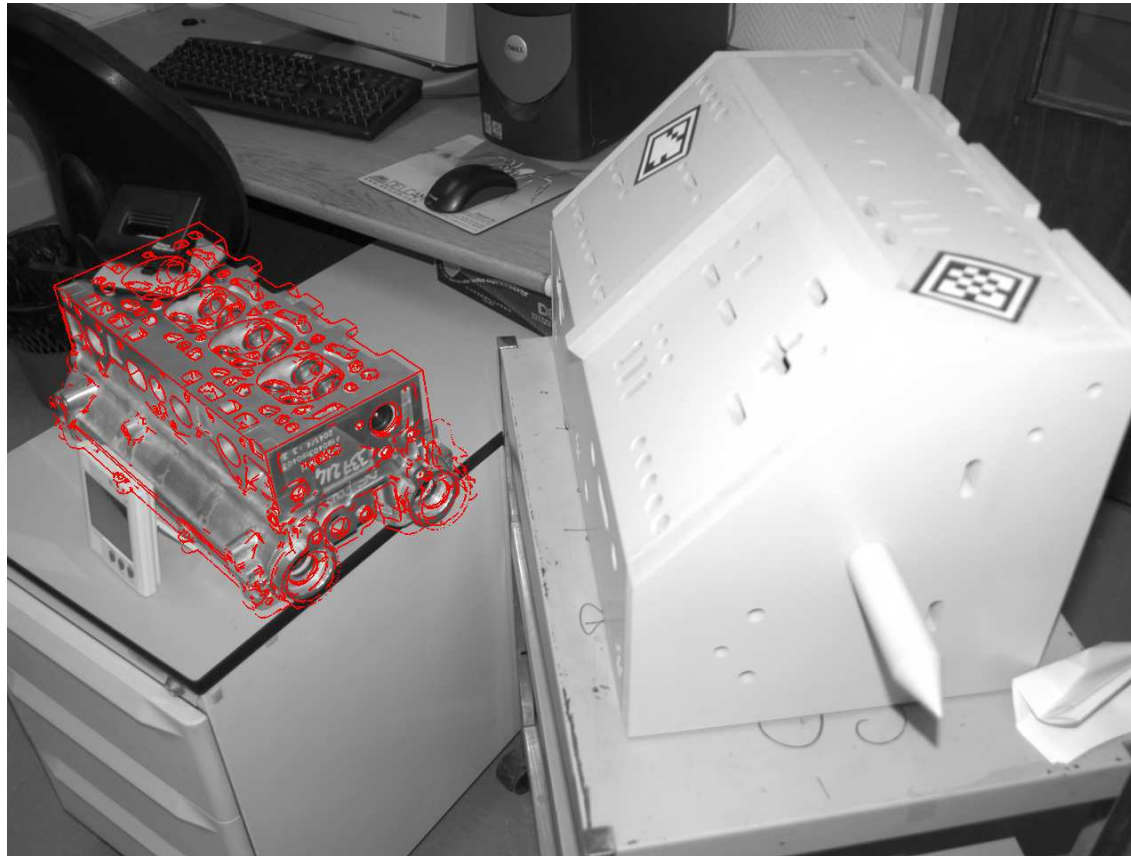


FIG. 6.39 – Exemple de recalage de la culasse selon l'approche 2D/2D dans un environnement non contrôlé.

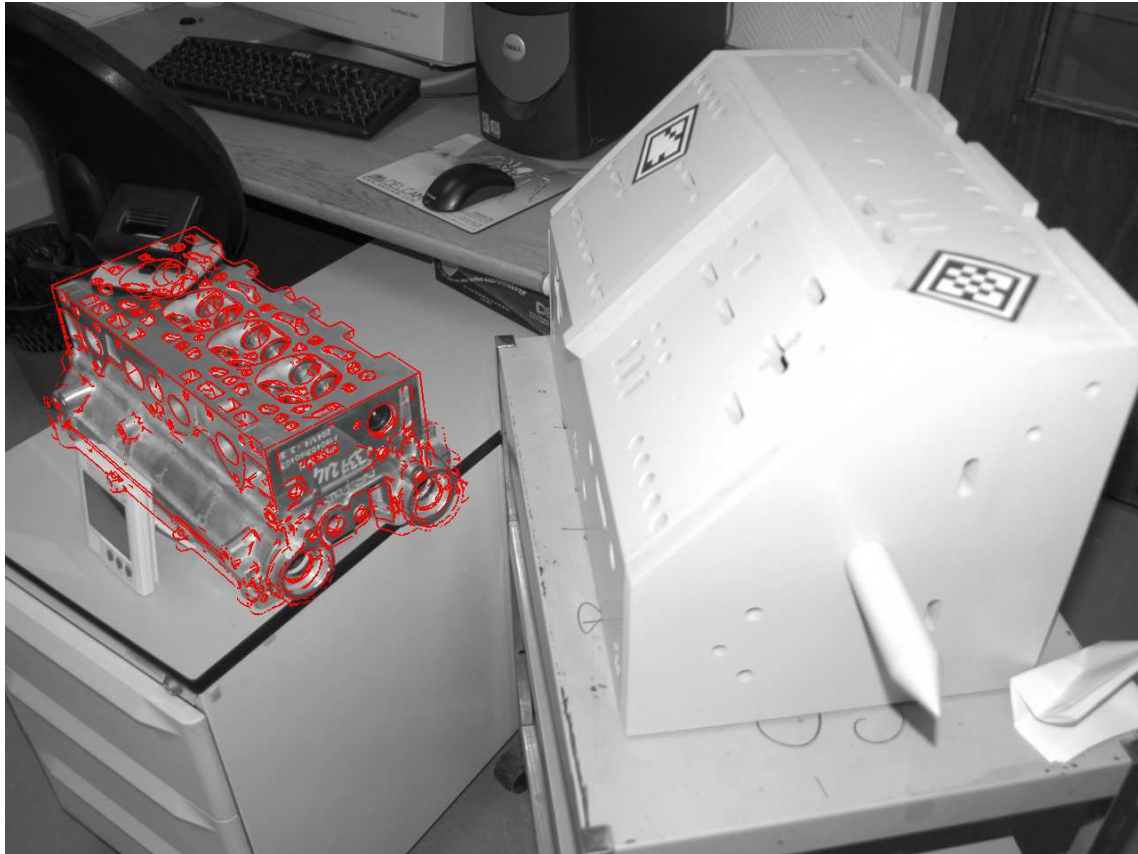


FIG. 6.40 – Exemple de recalage de la culasse selon l'approche 3D/2D dans un environnement non contrôlé.

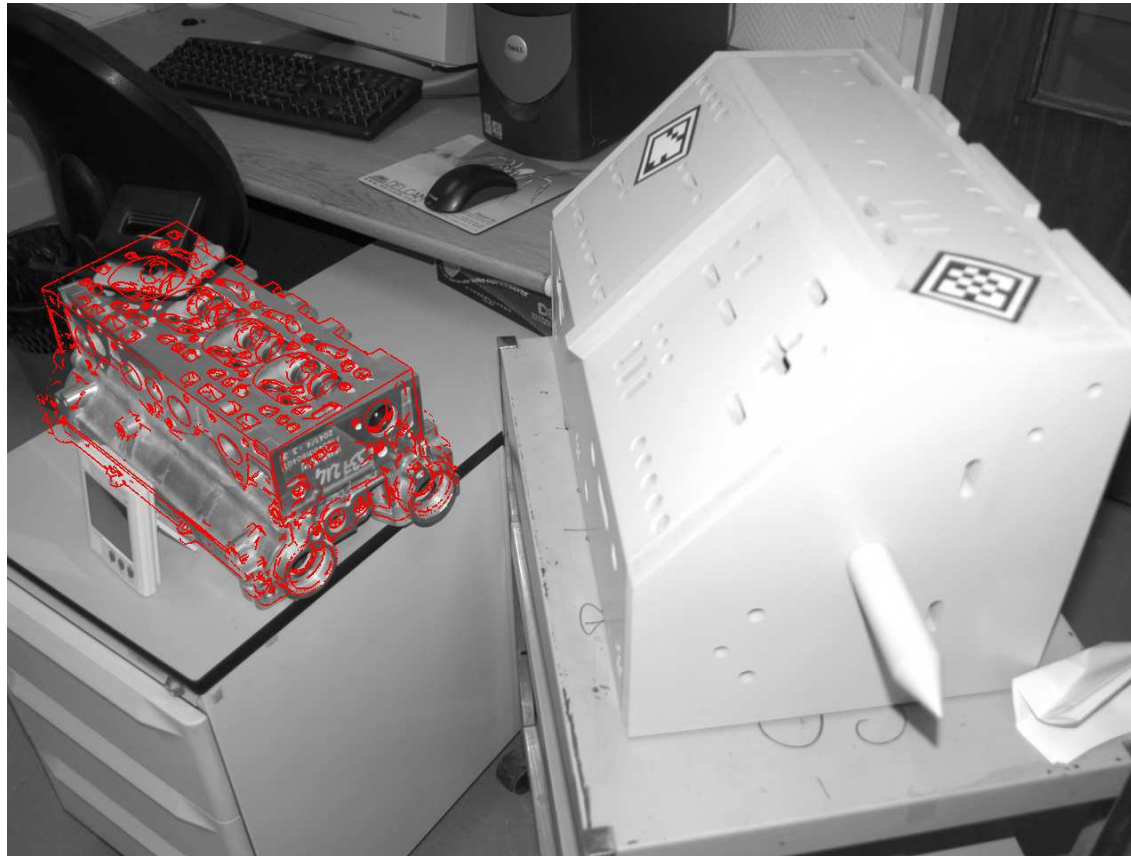


FIG. 6.41 – Exemple de recalage de la culasse selon l’approche 3D/3D dans un environnement non contrôlé.

6.5 Conclusion

En premier lieu, cette évaluation a permis de déterminer l'influence du choix des critères 2D ou 3D. Le fait que les méthodes 2D/2D et 3D/2D partagent des résultats très similaires sur l'ensemble des expériences indique que le critère 2D comme le critère 3D offrent tous deux un pouvoir discriminant équivalent lors de l'étape de filtrage géométrique des appariements. Ce résultat confirme donc les conclusions établies à l'issue de l'évaluation du processus de mise en correspondance géométrique. D'autre part, le très faible taux d'échec de ces deux approches (0 à 6%) indique que ce pouvoir discriminant permet de filtrer suffisamment les appariements pour qu'une pose précise puisse être estimée.

Par contre, le fait que la méthode 3D/3D présente des résultats en retrait face à ceux de la méthode 3D/2D indique que le critère 3D est peu adapté pour l'étape de calcul de pose. Lors de cette étape, il est primordial que le critère identifie la totalité des faux appariements. Or le critère 3D est évalué sur des estimations approximatives des poses des primitives de la scène. Par conséquent, l'erreur 3D associée à un appariement correct n'est pas négligeable. Pour parvenir à identifier les appariements corrects, le seuil sur le critère 3D ne peut donc pas être trop sévère. De ce fait, des appariements erronés peuvent être identifiés comme corrects, ces derniers perturbant le calcul de pose. La pose estimée est alors trop approximative pour que le processus de raffinement puisse converger vers la pose réelle de l'objet.

A l'opposé, le critère 2D repose sur la position 2D du centre des primitives, position dont l'évaluation est connue de manière relativement précise. Le critère 2D peut donc être sévère sans risquer de manquer trop d'appariements corrects. Il devient alors peu probable qu'un appariement erroné soit identifié comme correct. En l'absence de faux appariements, la pose estimée est généralement suffisamment proche de la pose réelle de l'objet pour que le processus de raffinement converge vers celle-ci.

D'autre part, cette évaluation a permis d'évaluer les performances du recalage. La première expérience a permis d'évaluer la robustesse de celui-ci aux changements de point de vues. La seconde expérience a montré que les méthodes 2D/2D et 3D/2D permettent généralement d'obtenir un recalage précis à partir d'un modèle léger, i.e. basé sur peu de vues. Enfin, la dernière expérience a montré la robustesse du recalage vis à vis de l'environnement extérieur : variations d'illuminations, présence d'occultations, présence d'un environnement chargé contenant de nombreux objets parasites, calibrage faible de la caméra, ...

Chapitre 7

Conclusion et Perspectives

Le travail présenté dans cette thèse a permis de développer un processus de recalage visuel. L'originalité de ce travail réside avant tout dans la nature des objets traités : il s'agit de pièces mécaniques complexes telles qu'une culasse de moteur ou d'une portière automobile. La difficulté intrinsèque à ce type d'objet vient de leur forme complexe, de leur nature métallique qui implique la présence de reflets spéculaires et l'absence de texture, ainsi que de la présence de structures répétitives à leur surface. D'autre part, le contexte de ces travaux implique l'absence de contrôle de l'environnement : l'éclairage n'est pas contrôlé, l'environnement peut être encombré, des occultations partielles peuvent survenir, et aucun a priori sur la pose de l'objet ne peut être exploité. Enfin, les applications visées (préhension par un robot, réalité augmentée, métrologie par vision) nécessitent un recalage précis. Le processus de recalage proposé repose avant tout sur l'élaboration d'un modèle caractérisant à la fois l'apparence et la géométrie de l'objet, ainsi qu'un processus de mise en correspondance permettant de gérer efficacement les nombreux faux appariements dus à la présence de motifs répétitifs. Ces deux points constituent les contributions majeures de notre travail. Les sections suivantes présentent un résumé de ce travail ainsi qu'un certain nombre de perspectives

7.1 Un modèle caractérisant apparence et géométrie 3D

La méthode proposée repose sur un unique modèle caractérisant à la fois l'apparence et la géométrie 3D de l'objet. Il s'agit d'un point essentiel puisque c'est cette caractérisation mixte qui permet d'obtenir un processus de mise en correspondance discriminant ainsi qu'une estimation précise de la pose. L'élaboration

de ce modèle nécessite la connaissance du modèle CAO de l'objet ainsi que de quelques vues de cet objet recalées avec le modèle CAO.

Ce modèle est établi à partir de primitives visuelles, issues des vues d'apprentissage, qui sont organisées dans un espace 3D. Plus précisément, les primitives visuelles employées correspondent à des descripteurs locaux caractérisant le voisinage d'un contour fermé. L'usage de descripteurs locaux permet d'obtenir une description de l'apparence de l'objet qui soit robuste aux occultations alors que l'usage de contours fermés pour établir les régions supports permet d'obtenir des descripteurs invariants aux transformations affines. Les objets étant peu texturés, les descripteurs locaux employés caractérisent, non pas la texture de la région support mais les contours présents dans celle-ci. Chacun de ces descripteurs se voit attribuer une pose, celle-ci correspondant à la position et à l'orientation 3D de la surface de l'objet se projetant sur la région support. De plus, chacun de ces descripteurs est enrichi par une description géométrique de cette surface qui lui est sous-jacente, le *motif 2D/3D*, ceci afin de pouvoir calculer une pose approximative de l'objet à partir d'un seul appariement.

7.2 Un processus de mise en correspondance discriminant

Un processus de mise en correspondance reposant sur l'apparence seule des primitives a été proposé. Cependant, une évaluation de celui-ci a permis de souligner l'impact des motifs répétitifs sur ses performances. La conclusion de cette évaluation a été que, du fait que les objets considérés dans le contexte de nos travaux comportent de nombreux motifs répétitifs, une mise en correspondance selon la seule apparence des primitives ne peut assurer un taux de faux appariements suffisamment faible pour que le calcul d'une pose en un temps raisonnable puisse être envisagé. Une étape supplémentaire consistant à filtrer les appariements a été ajoutée. Celle-ci exploite les contraintes de rigidité reliant les primitives du modèle, ainsi que la capacité à estimer une pose approximative de l'objet à partir d'un seul appariement. Deux variantes ont été proposées pour cette étape : la première travaille dans l'espace image alors que la seconde travaille dans l'espace objet.

Une évaluation a révélé que ces deux approches offrent des performances comparables, toutes deux permettant de réduire considérablement la quantité de faux appariements tout en conservant la quasi-totalité des appariements corrects. Néanmoins, l'approche travaillant dans l'espace objet présente l'avantage de ne dépendre que de paramètres dont la valeur peut être fixée sans avoir à connaître

a priori la distance d'observation. La différence entre les deux approches se situe donc plus au niveau de la simplicité de la mise en oeuvre de la solution par l'utilisateur final que des performances.

7.3 Un recalage visuel précis et robuste

Le processus de recalage visuel complet, c'est à dire incluant la mise en correspondance et le calcul de la pose, a été évalué sur des séquences réelles d'objets mécaniques complexes, tels qu'une culasse de moteur ou une portière de voiture. Cette évaluation a souligné la précision visuelle du recalage, les points de surface de l'objet se projetant généralement avec une précision de l'ordre d'un pixel d'erreur. Une telle précision est généralement suffisante pour des applications de réalité augmentée. En terme d'erreur 3D, l'évaluation a révélé des erreurs de l'ordre de quelques millimètres au niveau de la position des points de surface visibles de l'objet. Cette erreur correspond principalement à une translation le long de l'axe optique. Un recalage avec une telle précision est suffisant pour de nombreuses applications, comme la préhension d'une pièce par un robot. Si celle-ci reste insuffisante pour des applications de métrologie fine, la pose estimée est suffisante pour initialiser un processus de raffinement. Enfin, une expérience a permis d'illustrer la robustesse du processus aux conditions extérieures, c'est à dire en l'absence de contrôle de l'illumination, en présence d'occultations partielles, en présence d'un environnement chargé ou d'une caméra mal calibrée, ou encore lors d'importants changements d'échelles.

7.4 Perspectives

Si les évaluations expérimentales sont satisfaisantes, le processus proposé peut être à la fois amélioré, tant en terme de performances que de champs d'applications.

7.4.1 Amélioration du modèle

Différentes améliorations peuvent être apportées au modèle.

Lors de l'élaboration du modèle, la connaissance de la géométrie sous-jacente aux régions supports est sous-exploitée. Par exemple, il est possible de déterminer, à partir de celles-ci, les zones d'une région support qui n'appartiennent pas à l'objet. Dans le cas de la signature de *Shape Context*, l'information encodée dans

les secteurs correspondant à ces zones n'est pas pertinente puisque celle-ci ne caractérise pas l'objet mais l'arrière plan. Ces secteurs doivent donc être ignorés lors du calcul de distance entre cette signature du modèle et une signature issue d'un descripteur de l'image. D'une manière plus générale, il est possible d'attribuer un poids à chaque secteur du *Shape Context* en fonction de la planarité de la surface de l'objet correspondant à ce secteur. La stabilité de la position et de l'orientation des contours des zones non-planes étant moindre, cette pondération permettrait de réduire leur influence lors d'un calcul de distance avec la signature d'un descripteur issu de l'image.

D'autre part, il serait intéressant d'élaborer le modèle à partir de vues de synthèses, c'est à dire de vues générées à partir du modèle CAO de l'objet, et non de vues réelles de celui-ci. En effet, le modèle actuel nécessite de disposer de vues réelles en plus du modèle CAO, ces vues devant être finement recalées avec ce modèle CAO. L'acquisition et le recalage de ces vues est un processus complexe qui nuit à la simplicité de la mise en oeuvre de notre solution. L'usage de vues de synthèse permettrait non seulement de supprimer cette étape et donc de simplifier l'exploitation de notre solution, mais aussi de choisir de manière optimale les points de vue pour élaborer le modèle. De plus, si un modèle est élaboré avec trop peu de vues, il serait aussi plus facile de générer de nouvelles vues de synthèses pour compléter le modèle, ce qui n'est généralement pas le cas avec des vues réelles.

7.4.2 Amélioration du processus de mise en correspondance

Une première optimisation possible du processus de mise en correspondance correspond à l'optimisation, en terme de temps de calcul, de l'étape de mise en correspondance selon l'apparence. Des algorithmes de recherches rapides des plus proches voisins peuvent être utilisés pour accélérer ce processus [2]. Des structures de données telles que les *kd-trees* peuvent être employées pour organiser les primitives du modèle et faciliter cette recherche. Récemment, Lepetit et al. [47] ont proposé un processus de mise en correspondance, reposant sur des *randomized-trees*, dont les performances permettent son emploi dans un cadre temps-réel.

D'autres optimisations peuvent être apportées au processus de filtrage géométrique des appariements. En particulier, le critère de cohérence 3D/3D présenté dans la section 4.1.3 peut être amélioré à l'aide de la connaissance de l'incertitude sur les poses des primitives. En effet, le critère 3D/3D présente l'inconvénient de reposer sur l'estimation des poses des descripteurs de l'image, cette estimation

étant moins précise que l'estimation du centre des régions supports utilisée dans le critère 3D/2D. Par conséquent, prendre en compte l'incertitude sur les différents paramètres des poses, lors du calcul de la distance des trièdres, permettrait, par exemple, de réduire le poids de l'erreur de position le long de l'axe optique (la position le long de l'axe optique est généralement associée à une incertitude importante du fait qu'elle ne peut être aussi bien estimée que les deux autres composantes de position). Lorsque cette incertitude serait très grande, l'erreur le long de l'axe optique deviendrait négligeable et le critère 3D/3D tendrait alors vers un critère 3D/2D. La prise en compte de l'incertitude permettrait, en quelque sorte, de passer de manière continue d'un critère 3D/3D à un critère 3D/2D en fonction de la pertinence des poses des descripteurs considérés.

Enfin, dans le cas du critère 3D, il est envisageable d'associer un seuil spécifique à chaque descripteur du modèle. En effet, dans le cas de motifs répétitifs, il est souhaitable de disposer d'un seuil sur le critère de cohérence permettant d'identifier les seuls appariements corrects. Pour cela, il est nécessaire que ce seuil soit inférieur à la distance entre un descripteur et son plus proche descripteur équivalent. Or, cette distance n'est pas fixe d'un descripteur à l'autre. Une solution n'utilisant qu'un seul seuil risque donc de fixer celui-ci trop bas pour certains descripteurs et trop haut pour d'autres. Associer un seuil spécifique à chaque descripteur du modèle permettrait alors de répondre à ce problème. Celui-ci peut être établi de manière automatique, hors-ligne, permettant ainsi d'obtenir une plus grande simplicité de mise en oeuvre de la solution.

7.4.3 Amélioration du calcul de pose

La pose obtenue à l'issue du recalage est généralement suffisamment précise pour initialiser une nouvelle étape de raffinement de la pose. La solution envisagée est similaire à celle proposée par Vacchetti et al.[82]. Il s'agit alors d'ajuster les paramètres de la pose afin de minimiser la distance entre les contours de l'image et la reprojection des arrêtes du modèle CAO dans l'image.

D'autre part, l'intégration du processus de recalage dans une solution multi-caméra est aussi une piste envisagée pour accroître la précision des poses estimées ainsi que la robustesse du processus de recalage.

7.4.4 Extension du domaine d'application

Tout d'abord, le processus de recalage peut être généralisé à d'autres catégories d'objets. Il suffit pour cela de choisir des descripteurs locaux adaptés à la nature des objets.

Enfin, la solution proposée peut aisément être étendue pour réaliser de la reconnaissance de plusieurs objets. Il s'agit alors d'ajouter une étape permettant de gérer la présence, dans une même scène, de descripteurs correspondant à plusieurs objets connus. Il est notamment possible de s'inspirer des travaux de Rothganger et al. [71] sur ce point.

Bibliographie

- [1] N. Allezard, M. Dhome, and F. Jurie. Recognition of 3d textured objects by mixing view-based and model-based representations. In *International Conference of Pattern Recognition*, pages 960–963, Barcelona, Espagne, 2000.
- [2] S. Arya, D. Mount, N. Netanyahu, R. Silverman, and A. Wu. An optimal algorithm for approximate nearest neighbor searching. *Journal of the ACM*, 45 :891–923, 1998.
- [3] H. Asada and M. Brady. The curvature primal sketch. In *2nd Workshop Computer Vision : Representation and Control*, 1984.
- [4] D. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13 :111–122, 1981.
- [5] J. Belhumeur, P. Hespanha and D. Kriegman. Eigenfaces vs. fisherfaces : Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 :711–720, 1997.
- [6] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. Technical Report UCB//CSD00, UC Berkeley, Jan. 2001.
- [7] P. Besl and R. Jain. Three-dimensional object recognition. *ACM Computing Surveys*, 17 :75–145, 1985.
- [8] F. Bourgeois and J-C. Lassalle. An extension of the munkres algorithm for the assignment problem to rectangular matrices. *Commun. ACM*, 14(12) :802–804, 1971.
- [9] S. Bourgeois, S. Naudet-Colette, and M. Dhome. Cad model visual registration from closed-contour neighborhood descriptors. In *International Conference on Image Analysis and Recognition*, Septembre 2006.
- [10] S. Bourgeois, S. Naudet-Colette, and M. Dhome. Coarse visual registration from closed-contour neighborhood descriptor. In *International Conference of Pattern Recognition*, pages 283–287, Hong-Kong, Chine, Aout 2006.

- [11] S. Bourgeois, S. Naudet-Colette, and M. Dhome. Recalage d'un modèle cao à partir de descripteurs locaux de contours. In *15eme congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle*, Janvier 2006.
- [12] S. Bourgeois, S. Naudet-Colette, and M. Dhome. Visual registration of industrial metallic object from local contour descriptors. In *International Conference on Computer Graphics, Imaging and Visualization*, pages 273–280, Juillet 2006.
- [13] T. Breuel. *Geometric Aspects of Visual Object Recognition*. PhD thesis, Massachusetts Institute of Technology, 1992.
- [14] R. Brooks. Model-based three-dimensional interpretations of two-dimensional images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(2) :140–150, 1983.
- [15] J. Burns, A. Hanson, and E. Riseman. Extracting straight lines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4 :425–455, 1986.
- [16] J. Burns, R. Weiss, and E. Riseman. View variation of point-set and line-segment features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(1) :51–68, Janvier 1993.
- [17] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), 86.
- [18] T. Cass. Polynomial-time object recognition in the presence of clutter, occlusion, and uncertainty. In *DARPA Image Understanding Workshop*, 1992.
- [19] R. Chin and C. Dyer. Model-based recognition in robot vision. *ACM Computing Surveys*, 18 :67 – 108, 1986.
- [20] D. DeMenthon. Recognition and tracking of 3d objects by 1d search. In *DARPA93*, pages 653–659, 1993.
- [21] D. DeMenthon and L.S. Davis. Model-based object pose in 25 lines of code. *International Journal of Computer Vision*, 14 :123–141, 1995.
- [22] R. Deriche. Using canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*, 1 :167–187, 1987.
- [23] M. Dhome, M. Richetin, J.T. Lapreste, and G. Rives. Determination of the attitude of 3-d objects from a single perspective view. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(11) :1265–1278, 1989.
- [24] S. Dickinson, A. Pentland, and A. Rosenfeld. 3-d shape recovery using distributed aspect matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2) :174–198, 1992.

- [25] T. Drummond and R. Cipolla. Real-time visual tracking of complex structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7) :932–946, 2002.
- [26] M.A. Fischle and R.C. Bolles. Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography. *Graphics and Image Processing*, 24(6) :381–395, June 1981.
- [27] M Fischler and R. Bolles. Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6) :381–395, 1981.
- [28] J. Flusser and T. Suk. Pattern recognition by affine moment invariants. *Pattern Recognition*, 26 :167–174, 1993.
- [29] Z. Gigus and J. Malik. Computing the aspect graph for line drawings of polyhedral objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2 :113–122, 1990.
- [30] W. Grimson. The combinatorics of heuristic search term for object recognition in cluttered environment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9) :920–935, 1991.
- [31] W. Grimson. Object recognition by computer : The role of geometric constraints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3) :255–274, 90.
- [32] C. Harris. *Active Vision*, chapter Tracking with rigid models., pages 59–73. MIT Press, 1992.
- [33] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, 1988.
- [34] R. Horaud, B. Conio, and O. Le Boulleux. An analytical solution for the perspective-4-point problem. *Computer Vision, Graphics, and Image Processing*, 47 :33–44, 1989.
- [35] C. Huang, A. Camps, and T. Kanungo. Object recognition using appearance-based parts and relations. In *International Joint Conferences on Artificial Intelligence*, 1997.
- [36] X. Huang, B. Wang, and L. Zhang. A new scheme for extraction of affine invariant descriptor and affine motion estimation based on independent component analysis. *Pattern Recognition Letters*, 26 :1244–1255, 2005.
- [37] Z. Huang and F. Cohen. Affine-invariant b-spline moment for curve matching. *IEEE. Trans. Image Processing*, 5(10) :1473–1480, 1996.

- [38] D. Huttenlocher and S. Ullman. Recognizing solid objects by alignment. In *DARPA Image Understanding Workshop*, volume 2, pages 1114–1122, 1988.
- [39] D.P. Huttenlocher and S. Ullman. Object recognition using alignment. In *First International Conference on Computer Vision*, 1987.
- [40] F. Jurie. Solution of the simultaneous pose and correspondence problem using gaussian error model. *Computer Vision and Image Understanding*, 73(3) :357–373, 1999.
- [41] T. Kadir. *Scale, Saliency and Scene Description*. PhD thesis, University of Oxford, 2002.
- [42] M. Khalil and M. Bayoumi. Affine invariants for object recognition using the wavelet transform. *Pattern Recognition Letters*, 23 :57–72, 2002.
- [43] Y. Lamdan and H. Wolfson. Geometric hashing : A general and efficient model-based recognition scheme. Technical Report 368, New York University, Robotics Research Laboratory, Department of Computer Science, 1998.
- [44] S. Lazebnik. *Local, Semi-Local and Global Models for Texture, Object and Scene Recognition*. PhD thesis, University of Illinois at Urbana Champaign, 2006.
- [45] D. Lee and S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401 :788–7991, 1999.
- [46] A. Leonardis and H. Bischof. Robust recognition using eigenimages. *Computer Vision and Image Understanding*, 78(1) :99–118, 2000.
- [47] V. Lepetit and P. Fua. Keypoint recognition using randomized trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9) :1465–1479, 2006.
- [48] S. Li, X. Hou, and H. Zhang. Learning spatially localized, part-based representation. In *Conference on Computer Vision and Pattern Recognition*, Hawaii, 2001.
- [49] T. Lindeberg. *Discrete Scale-Space Theory and the Scale-Space Primal Sketch*. PhD thesis, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, S-100 44 Stockholm, Suède, Mai 1991.
- [50] D.G. Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31(3) :355–395, March 1987.
- [51] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2) :911–110, 2004.

- [52] D. Macrini, A. Shokoufandeh, D. Dickinson, and K. Siddiqi. View-based 3-d object recognition using shock graphs. In *International Conference of Pattern Recognition*, Quebec, Canada, 2002.
- [53] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robustwide baseline stereo from maximally stable extremal regions. In *The British Machine Vision Conference*, Cardiff, UK, 2002.
- [54] A.I. Medalia. Dynamic shape factors of particles. *Powder Technology*, (4) :117–138, 1970.
- [55] G. Medioni and A. Francois. 3-d structures for generic object recognition. In *International Conference of Pattern Recognition*, Barcelona, Espagne, 2000.
- [56] K. Mikolajczyk. *Detection of local features invariant to affine transformations*,. PhD thesis, INPG Grenoble, 2002.
- [57] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10) :1615–1630, 2005.
- [58] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schafalitzky, T. Kadir, and L. Van Goo. A comparison of affine region detectors. *International Journal of Computer Vision*, 65, 2005.
- [59] K. Mikolajczyk, A. Zisserman, and C. Schmid. Shape recognition with edge-based features. In *The British Machine Vision Conference*, Norwich, UK., 2003.
- [60] F. Mokhtarian. Silhouette-based occluded object recognition through curvature scale space. *Machine Vision and Applications*, 10(3) :87–97, 1997.
- [61] H. Murase and S. Nayar. Visual learning and recognition of 3-d objects from appearance. *International Journal of Computer Vision*, 14 :5–24, 1995.
- [62] R. Nevatia and T.O. Binford. Description and recognition of curved objects. *Artificial Intelligence*, 8 :77–98, 1977.
- [63] S. Obdrzalek and J. Matas. Local affine frames for image retrieval. In *International Conference The Challenge of Image and Video Retrieval*, volume 1, pages 318–327, 2002.
- [64] S. Obdrzalek and J. Matas. Object recognition using local affine frames on distinguished regions. In *The British Machine Vision Conference*, Cardiff, UK, 2002.
- [65] A. Oirrak, M. Daoudi, and D. Aboutajdine. Affine invariant descriptors using fourier series. *Pattern Recognition Letters*, 23 :1109–1118, 2002.

- [66] S. Petitjean. *Géométrie énumérative et contacts de variétés linéaires : application aux graphes d'aspects d'objets courbes*. PhD thesis, Institut National Polytechnique de Lorraine, 1995.
- [67] J. Ponce, D. Chelberg, and W. Mann. Invariant properties of straight homogeneous generalized cylinders and their contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(9) :951–966, Septembre 1989.
- [68] A. Pope and D. Lowe. Probabilistic models of appearance for 3-d object recognition. *International Journal of Computer Vision*, 40(2) :149–167, 2000.
- [69] R. Pope and D. Lowe. Learning appearance models for object recognition. In *International Workshop on Object Representation for Computer Vision*, 1996.
- [70] L. G Roberts. Machine perception of three-dimensional solids. *Optical and Electro-Optical Information Processing*, pages 159–197, 1965.
- [71] F. Rothganger, S. Lazebnik, C. Schmid, and J. Ponce. 3d object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints. *International Journal of Computer Vision*, 66(3), 2006.
- [72] C.A. Rothwell, A. Zisserman, D.A. Forsyth, and J.L. Mundy. Planar object recognition using projective shape representation. In *International Journal of Computer Vision*, volume 16, pages 57–99, 1995.
- [73] R. D. Schiffenbauer. A survey of aspect graphs. Technical Report TR-CIS-2001-01, Polytechnic University of Brooklyn, 2001.
- [74] C. Schmid. *Appariement d'images par invariants locaux de niveaux de gris*. PhD thesis, Institut National Polytechnique de Grenoble, 1996.
- [75] S. Sclaroff and A. Pentland. odal matching for correspondence and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(6) :545–561, 1995.
- [76] T. Sebastian, P. Klein, and B. Kimia. Recognition of shapes by editing shock graphs. In *International Conference on Computer Vision*, Vancouver, Canada, 2001.
- [77] A. Selinger. *Analysis and Applications of Feature-Based Object Recognition*. PhD thesis, University of Rochester, 2001.
- [78] A. Sethi, D. Renaudi, D. Kreigman, and J. Ponce. Curve and surface duals and the recognition of curved 3d objects from their silhouette. *International Journal of Computer Vision*, 58(1) :73 – 86, Juin 2004.

- [79] A. Torralba, K. Murphy, W. Freeman, and M. Rubin. Context-based vision system for place and object recognition. In *International Conference on Computer Vision*, Nice, France., 2003.
- [80] Q. Tueng and W. Boles. Wavelet-based affine invariant representation : a tool for recognizing planar objects in 3d space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 :846–857, 1997.
- [81] M. Turk and A. Pentland. Eigenfaces for recognition. *Cognitive Neuroscience*, 3(1), 1991.
- [82] L. Vacchetti, V. Lepetit, and P. Fua. Combining edge and texture information for real-time accurate 3d camera tracking. In *International Symposium on Mixed and Augmented Reality*, 2004.
- [83] S. Zhang, R. Deriche, O. Faugeras, and Q. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Technical Report 2273, INRIA, 1994.
- [84] A. Zhao and J. Chen. Affine curve moment invariants for shape recognition. *Pattern Recognition*, 30(6) :895–901, 1997.
- [85] K. Zuiderveld. *Graphics Gems IV*, chapter Contrast Limited Adaptive Histogram Equalization, pages 474–485. Kaufmann, M., 1994.

Annexe A

Précision, Rappel et Courbe de Précision/Rappel

La *précision* et le *rappel* sont deux mesures permettant notamment d'évaluer un processus de mise en correspondance.

Comme l'illustre la figure A.1, l'espace des appariements contient deux ensembles : celui des appariements corrects et celui des appariements établis par le processus de mise en correspondance.

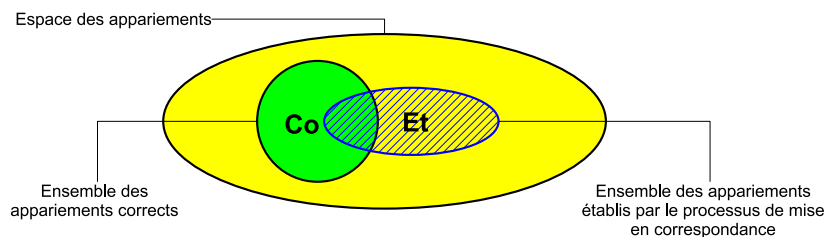


FIG. A.1 – L'espace des appariements

Ces deux ensembles n'étant généralement pas disjoints, ceux-ci définissent trois sous ensembles :

- A : l'ensemble des appariements corrects que le processus de mise en correspondance n'a pas identifié, $A = Co \setminus Et$
- B : l'ensemble des appariements corrects que le processus de mise en correspondance a identifié, $B = Co \cap Et$
- C : l'ensemble des faux appariements établis par le processus de mise en correspondance, $C = Et \setminus Co$.

La Précision La **précision** correspond au rapport entre le nombre d'appariements corrects identifiés par le processus de mise en correspondance et le nombre de d'appariements établis par le processus de mise en correspondance (cf figure A.2), soit :

$$\text{Précision} = \frac{\text{Card}(B)}{\text{Card}(B) + \text{Card}(C)}$$

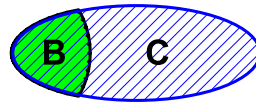
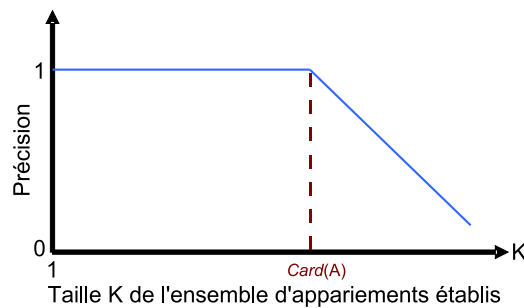


FIG. A.2 – Calcul de la précision

FIG. A.3 – Variation de la précision en fonction de la taille K de l'ensemble d'appariements établis dans le cas d'une sélection idéale.

Une précision est de 1 signifie que le processus de mise en correspondance n'a établi aucun faux appariement. A l'opposé, une précision de 0 correspond au cas où tous les appariements établis par le processus sont erronés. La taille de l'ensemble des appariements établis par un processus de mise en correspondance repose généralement sur un paramètre, ce dernier pouvant correspondre à la taille souhaitée de l'ensemble d'appariements ou à un seuil sur un critère de qualité des appariements. En modifiant ce paramètre, il est possible d'augmenter la taille de l'ensemble des appariements établi par le processus. On peut alors établir la courbe de la précision en fonction de ce paramètre. Dans le cas d'un processus de mise en correspondance idéal, cette courbe présente la forme caractéristique illustrée par la figure A.3. Dans un premier temps, la précision reste à 1 jusqu'à ce que la taille K de l'ensemble des appariements établis atteigne celle de l'ensemble

des appariements corrects contenus dans l'espace des appariements. Dès lors que cette taille est dépassée, le processus de mise en correspondance ne peut ajouter que des appariements erronés, ce qui entraîne la diminution de la précision.

Le Rappel Le **Rappel** correspond au rapport entre le nombre d'appariements corrects identifiés par le processus de mise en correspondance et le nombre total d'appariements corrects dans l'espace des appariements (cf. figure A.4), soit :

$$\text{Rappel} = \frac{\text{Card}(B)}{\text{Card}(A) + \text{Card}(B)}$$

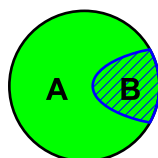


FIG. A.4 – Calcul du Rappel

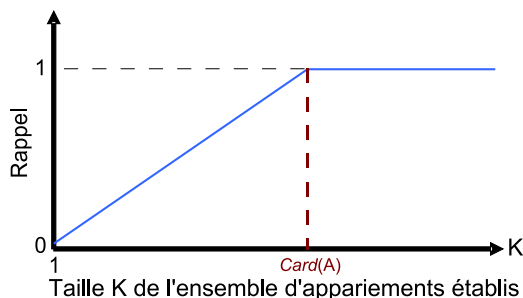


FIG. A.5 – Variation du rappel en fonction de la taille K de l'ensemble des appariements établis dans le cas d'une sélection idéale.

Un rappel de 1 signifie que le processus de mise en correspondance a identifié l'ensemble des appariements corrects de l'espace des appariements. A l'inverse, un rappel de 0 correspond au cas où le processus n'a établi que des faux appariements. Tout comme pour la précision, il est généralement possible d'établir la courbe du rappel, celle-ci correspondant aux variations du rappel en fonction du paramètre déterminant la taille de l'ensemble des appariements. Dans le cas d'un processus de mise en correspondance idéal, cette courbe présente la forme caractéristique illustrée par la figure A.5. Dans un premier temps, le rappel augmente

et atteint un rappel de 1 lorsque la taille de l'ensemble des appariements établis par le processus de mise en correspondance est égal à celle de l'ensemble des appariements corrects contenus dans l'espace des appariements. Dès lors que cette taille est atteinte, le processus de mise en correspondance ne peut ajouter que des appariements erronés, ce qui laisse alors le rappel inchangé.

La courbe de Précision/Rappel La courbe de précision et celle de rappel sont toutes deux définies par rapport à la taille t de l'ensemble des appariements établis par le processus de mise en correspondance. Il est donc possible d'établir une courbe 2D paramétrée ($\text{Précision}(t), \text{Rappel}(t)$). Cette courbe, dite de *Précision/-Rappel*, permet de mettre en évidence la relation liant les deux mesures. L'interprétation des résultats est alors simplifiée puisque celle-ci permet de déterminer la précision atteinte pour un rappel donné ou inversement.

La courbe de précision/rappel dans le cas d'un processus de mise en correspondance idéal est illustrée dans la figure A.6.

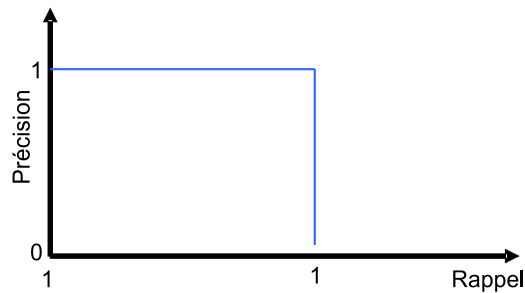


FIG. A.6 – Courbe de Précision/Rappel dans le cas d'une sélection idéale.

Annexe B

Distance des trièdres

La distance des trièdres permet d'évaluer l'amplitude d'un changement de repère $[\mathbf{R}|\mathbf{T}]$ au moyen d'une seule mesure, celle-ci intégrant à la fois la composante translationnelle \mathbf{T} et la composante rotationnelle \mathbf{R} du changement de repère. Afin de normaliser cette distance, deux paramètres, d_{max} et θ_{max} sont introduits. Le premier paramètre correspond à la norme des translations pures pour lesquelles la distance vaut 1. Le second correspond, dans le cas d'une rotation pure représentée sous la forme axe/angle, à l'angle de rotation pour lequel la distance prend la valeur 1. Ces deux paramètres sont à fixer par l'utilisateur.

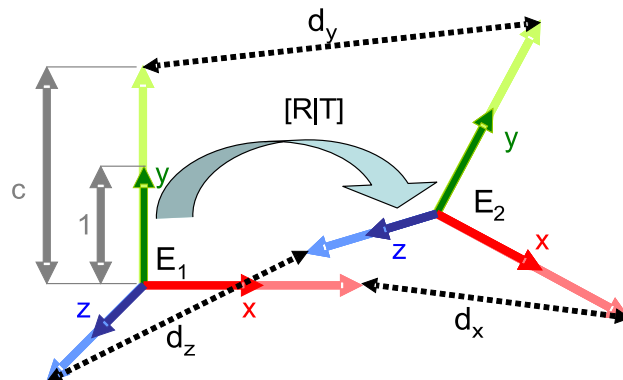


FIG. B.1 – Principe de l'erreur des trièdres. Les repères E_1 et E_2 sont reliés par une transformation rigide $[\mathbf{R}|\mathbf{T}]$. A chaque repère est associé un trièdre, orienté selon les axes du repère, et de côtés de longueur c . L'erreur des trièdres correspond à la distance moyenne entre les sommets des deux trièdres. Cette distance est normalisée par une distance d_{Max} .

Afin de présenter cette distance, notons E_1 et E_2 deux repères liés par une transformation $[R|T]$. La distance des trièdres entre ces deux repères est définie par :

$$ErrTriedre(E_1, E_2) = \frac{d_x^2 + d_y^2 + d_z^2}{\Delta}$$

où, comme l'illustre la figure B.1, les d_i correspondent aux distances entre les sommets des trièdres, de côtés de taille c , associés aux repères E_1 et E_2 ; et où Δ correspond à un facteur permettant de normaliser la distance en fonction des paramètres d_{max} et θ_{max} fixés. Le calcul de la valeur de c et de Δ est donné plus loin (cf. équations B et B.1). La figure B.2 illustre le principe de cette normalisation.

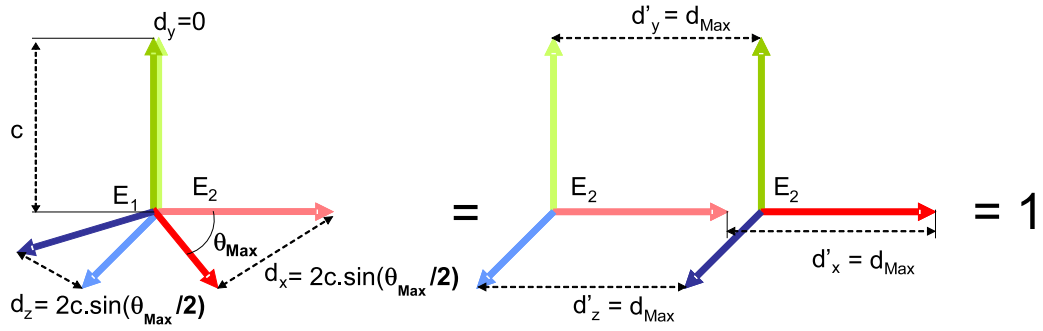


FIG. B.2 – Pondération des composantes rotationnelles et translationnelles dans la distance du trièdre. La taille des côtés des trièdres est établie à partir de l'angle θ_{Max} et la distance d_{max} , ceci de manière à ce que la distance entre deux repères reliés par une translation pure de norme d_{max} vaille 1 tout comme la distance entre deux repères reliés par une rotation pure d'angle θ_{max}

On notera que, quelque soit la valeur de c , la distance obtenue pour une translation pure T ne dépend que de la norme de celle-ci puisque $d_x = d_y = d_z = \|T\|$. Puisque la distance doit prendre la valeur 1 pour une translation pure de norme d_{max} , on fixe la valeur de Δ à :

$$\Delta = 3 \cdot d_{max}^2$$

D'autre part, on peut démontrer que la distance entre deux repères reliés par une rotation pure ne dépend pas de l'axe de rotation mais uniquement de l'angle. Pour cela, notons R une rotation d'angle θ autour de l'axe défini par le vecteur unitaire $v = [v_x, v_y, v_z]^T$. D'après la formule de Rodrigues, on a :

$$R = \begin{bmatrix} \cos(\theta) + v_x^2(1 - \cos(\theta)) & v_x v_y(1 - \cos(\theta)) - v_z \sin(\theta) & v_y \sin(\theta) + v_x v_z(1 - \cos(\theta)) \\ v_z \sin(\theta) + v_x v_y(1 - \cos(\theta)) & \cos(\theta) + v_y^2(1 - \cos(\theta)) & v_y v_z(1 - \cos(\theta)) - v_x \sin(\theta) \\ v_x v_z(1 - \cos(\theta)) - v_y \sin(\theta) & v_x \sin(\theta) + v_y v_z(1 - \cos(\theta)) & \cos(\theta) + v_z^2(1 - \cos(\theta)) \end{bmatrix}$$

On peut alors montrer que la valeur de $\delta = d_x^2 + d_y^2 + d_z^2$ ne dépend pas de v :

$$\begin{aligned}
\delta &= d_x^2 + d_y^2 + d_z^2 \\
&= \left\| R \begin{pmatrix} c \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} c \\ 0 \\ 0 \end{pmatrix} \right\|^2 + \left\| R \begin{pmatrix} 0 \\ c \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ c \\ 0 \end{pmatrix} \right\|^2 + \left\| R \begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix} \right\|^2 \\
&= 2c^2(v_z^2 \sin^2(\theta) + v_x^2 v_y^2 (1 - \cos(\theta))^2) + 2c^2(v_y^2 \sin^2(\theta) + v_x^2 v_z^2 (1 - \cos(\theta))^2) \\
&\quad + 2c^2(v_x^2 \sin^2(\theta) + v_y^2 v_z^2 (1 - \cos(\theta))^2) \\
&\quad + c^2(1 + v_x^2)^2(1 - \cos(\theta))^2 + c^2(1 + v_y^2)^2(1 - \cos(\theta))^2 \\
&\quad + c^2(1 + v_z^2)^2(1 - \cos(\theta))^2 \\
&= 2c^2 \sin^2(\theta)(v_x^2 + v_y^2 + v_z^2) + c^2(1 - \cos(\theta))^2(v_x^4 - 2v_x^2 + 1 + v_y^4 - 2v_y^2 + 1 \\
&\quad + v_z^4 - 2v_z^2 + 1 + 2v_x^2 v_y^2 + 2v_x^2 v_z^2 + 2v_y^2 v_z^2) \\
&= 2c^2 \sin^2(\theta)(v_x^2 + v_y^2 + v_z^2) + c^2(1 - \cos(\theta))^2((v_x^2 + v_y^2 + v_z^2)^2 - 2(v_x^2 + v_y^2 + v_z^2) + 3) \\
&= 2c^2 \sin^2(\theta) + 2c^2(1 - \cos(\theta))^2 \\
&= 8c^2 \sin^2(\theta/2)
\end{aligned}$$

Par conséquent, pour que la distance soit égale à 1 pour une rotation pure d'angle θ_{max} , il suffit de fixer la valeur de c de manière à ce que $\delta/(3d_{max}^2) = 1$, soit :

$$c = \sqrt{\frac{3d_{max}^2}{8 \sin^2(\theta_{max}/2)}} \quad (\text{B.1})$$

Annexe C

Résultats du recalage avec un modèle léger

Dans les pages suivantes, les résultats correspondant à l'expérience de recalage avec un modèle léger présentée dans la section 6.3 sont présentés pour la culasse et l'étalon. Les images de recalages sont accompagnées des représentations en fausses couleurs des erreurs de reprojections et d'erreur de position 3D des points. A défaut de pouvoir présenter ces images pour la totalité des vues de chaque objet, celles présentées correspondent à un échantillonnage régulier de l'ensemble des vues. Les résultats sur ces quelques vues offrent néanmoins une vision relativement représentative de la répartition des erreurs de recalage à la surface de l'objet. L'interprétation, en terme d'erreur, des fausses couleurs est donnée par la figure C.1 pour l'erreur de reprojection et par la figure C.2 pour l'erreur de position 3D des points de surface. Les résultats concernant l'étalon sont présentés à partir de la page 189 alors que ceux de la culasse sont présentés à partir de la page 211.

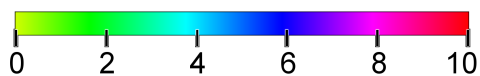


FIG. C.1 – Echelle utilisée pour représenter l'erreur de reprojection (exprimée en pixels) en fausses couleurs.

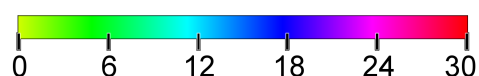


FIG. C.2 – Echelle utilisée pour représenter l'erreur de position 3D (exprimée en mm) des points de surface en fausses couleurs.

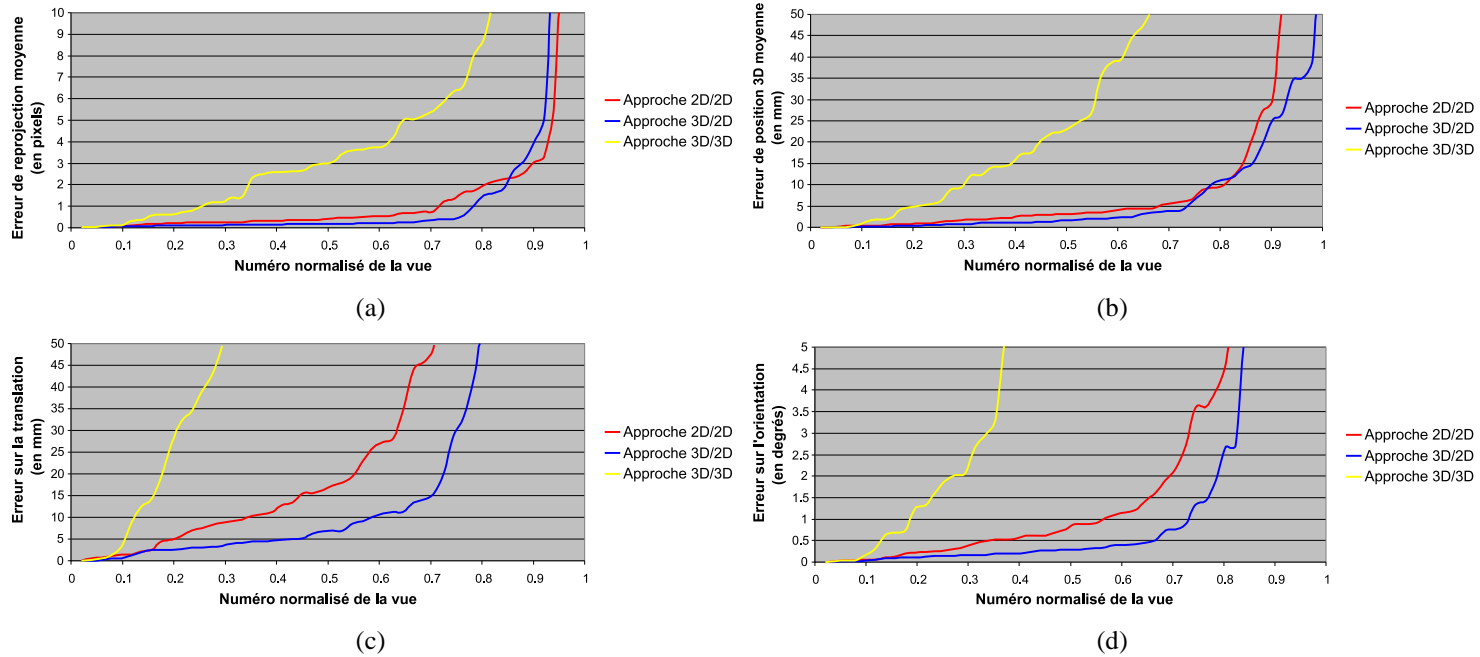


FIG. C.3 – Résultats de l'évaluation sur les 50 vues de l'étalon. Le graphe (a) représente l'erreur de reprojection pour chacune des 50 vues, alors que le graphe (b) représente l'erreur de position 3D des points de surface, et les graphes (c) et (d) représentent respectivement l'erreur de translation et d'orientation de la pose. Pour chacun des graphes, les vues sont triées par ordre croissant d'erreur. L'abscisse des graphes correspond au numéro normalisé de la vue (ie la vue ayant l'erreur minimale correspond au numéro 0, la vue ayant l'erreur maximale correspond au numéro 1 et la vue ayant l'erreur médiane correspond à l'abscisse 0.5).

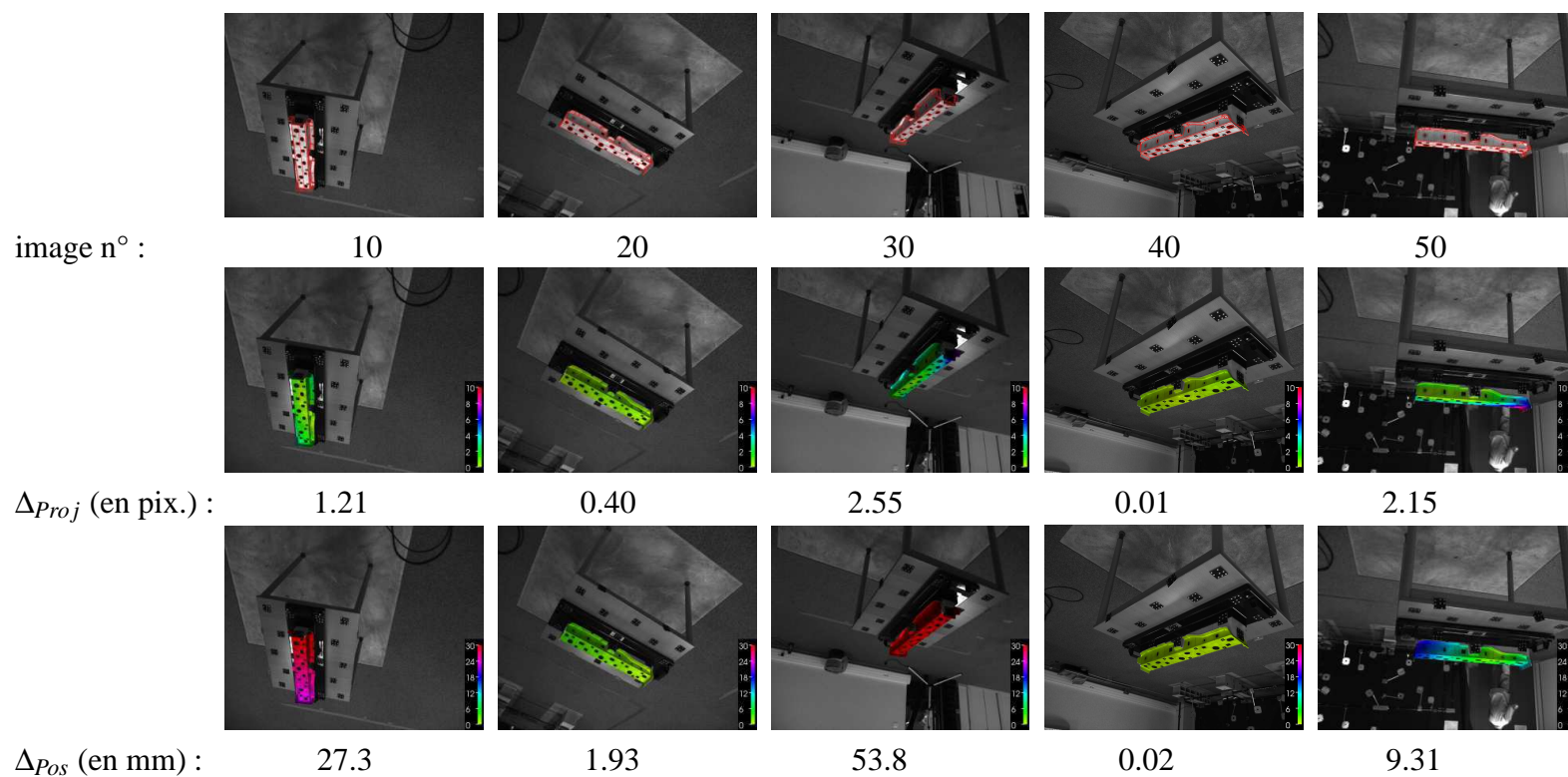


FIG. C.4 – Exemples de résultats pour l'étalon avec la méthode 2D/2D. 1ère ligne d'images : reprojection de l'étalon . 2ème ligne d'images : l'erreur de reprojection (en pixels) représentée en fausses couleurs. Δ_{proj} l'erreur de reprojection moyenne. 3ème ligne d'images : l'erreur de position 3D des points de surface (en mm) représentée en fausse couleurs. Δ_{pos} : l'erreur de position 3D moyenne.

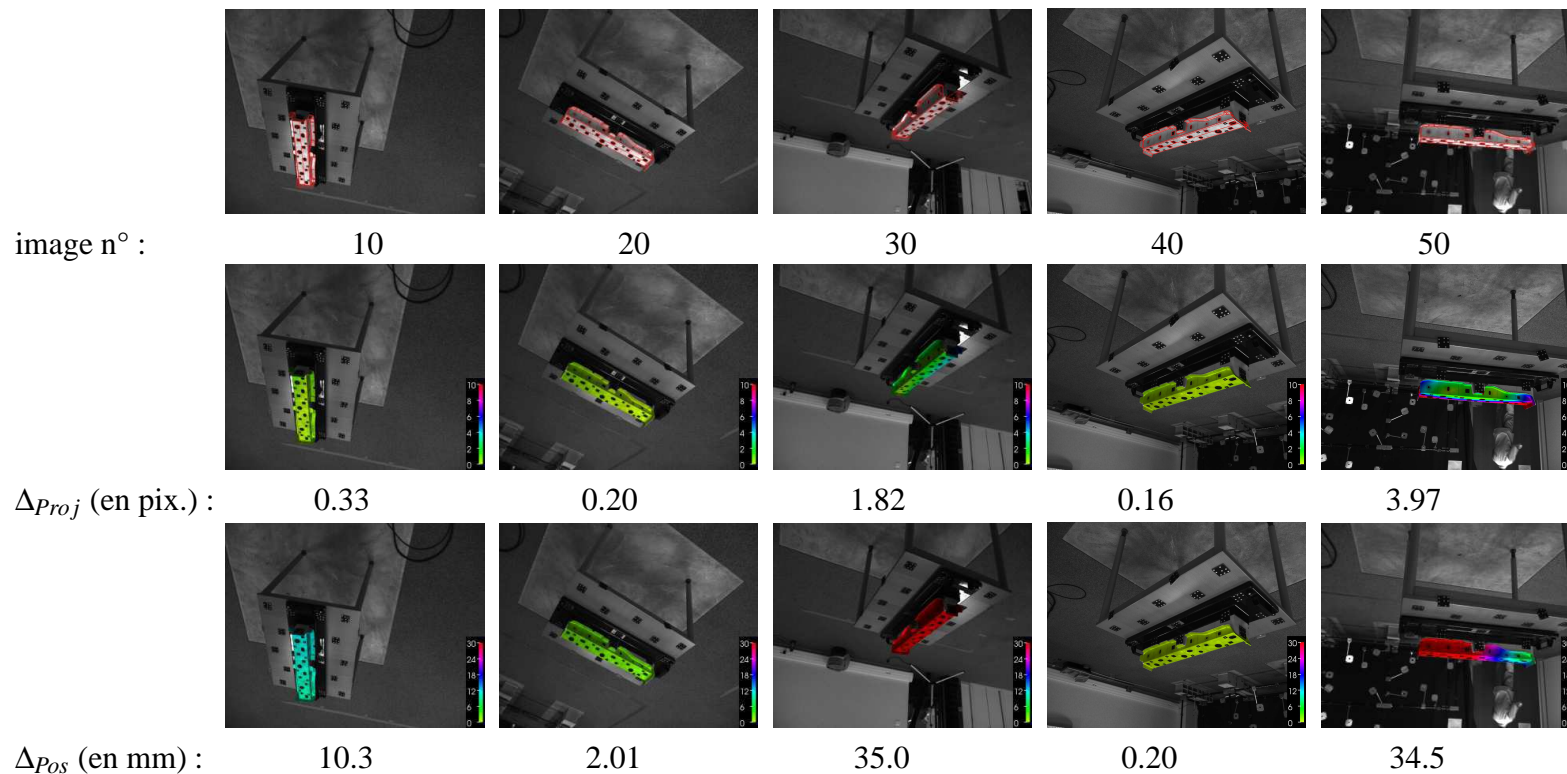


FIG. C.5 – Exemples de résultats pour l'étalon avec la méthode 3D/2D. 1ère ligne d'images : reprojection de l'étalon. 2ème ligne d'images : l'erreur de reprojection (en pixels) représentée en fausses couleurs. Δ_{proj} l'erreur de reprojection moyenne. 3ème ligne d'images : l'erreur de position 3D des points de surface (en mm) représentée en fausses couleurs. Δ_{pos} : l'erreur de position 3D moyenne.

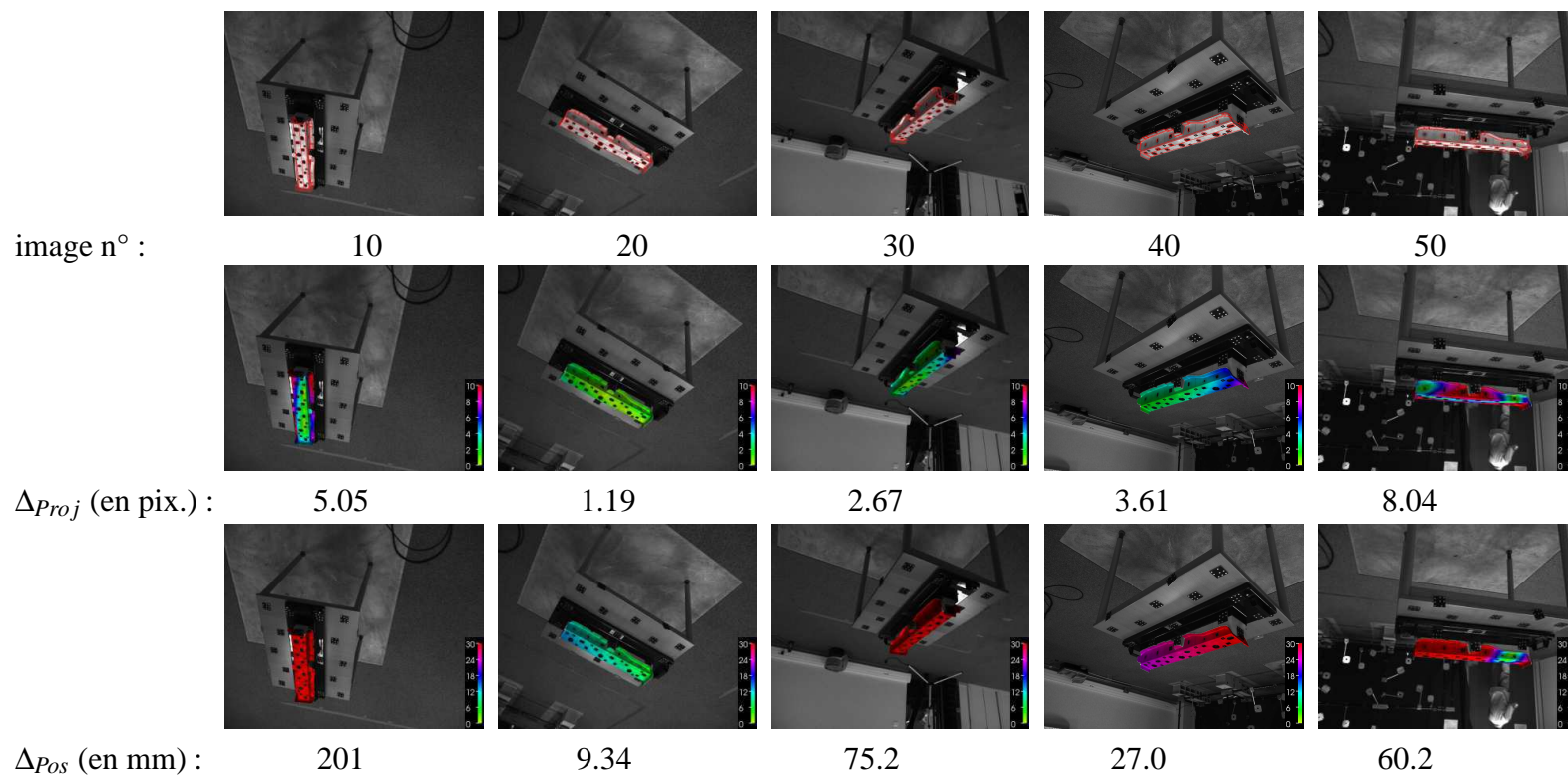
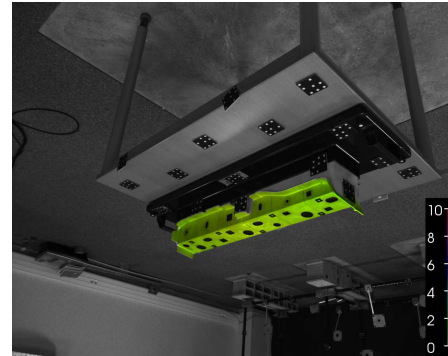


FIG. C.6 – Exemples de résultats pour l'étalon avec la méthode 3D/3D. 1^{ère} ligne d'images : reprojection de l'étalon . 2^{ème} ligne d'images : l'erreur de reprojection (en pixels) représentée en fausses couleurs. Δ_{proj} l'erreur de reprojection moyenne. 3^{ème} ligne d'images : l'erreur de position 3D des points de surface (en mm) représentée en fausses couleurs. Δ_{pos} : l'erreur de position 3D moyenne.

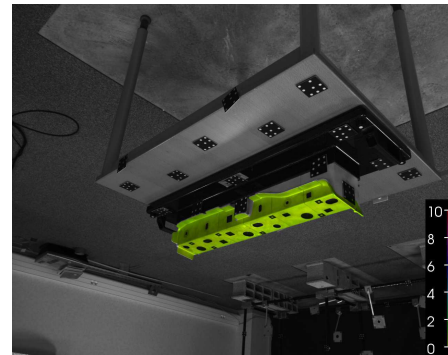
Reprojection du modèle



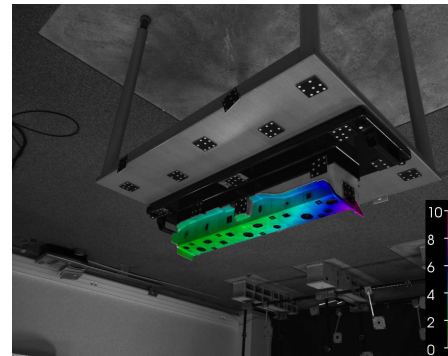
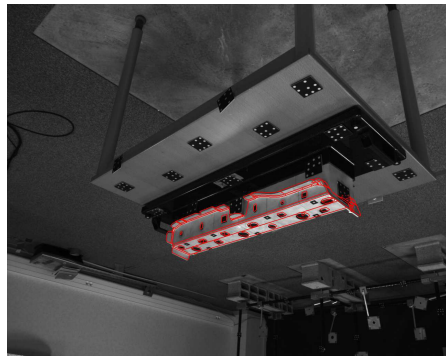
Erreur de reprojection



Méthode 2D/2D



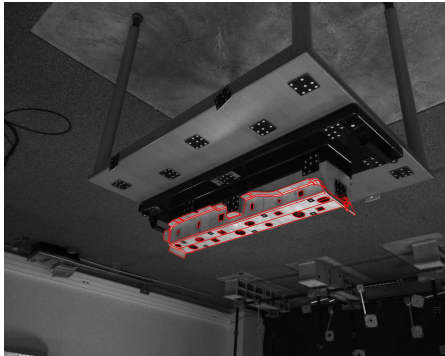
Méthode 3D/2D



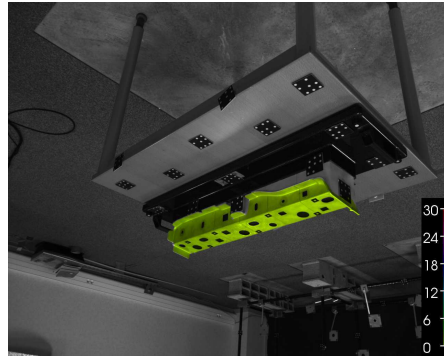
Méthode 3D/3D

FIG. C.7 – Erreur de reprojection du meilleur recalage de l'étalon. *Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de reprojection associée (en pixels).*

Reprojection du modèle



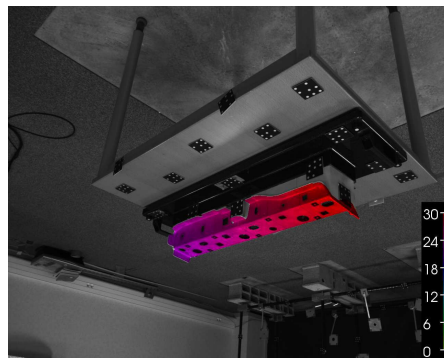
Erreur de position 3D



Méthode 2D/2D



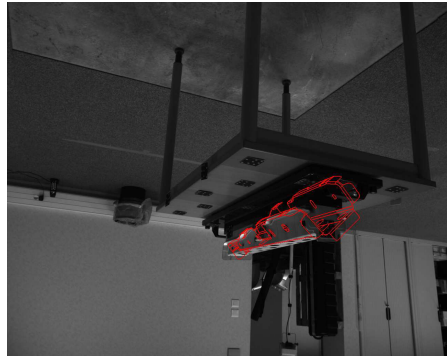
Méthode 3D/2D



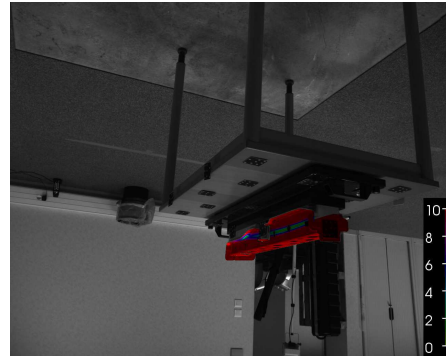
Méthode 3D/3D

FIG. C.8 – Erreur de position 3D des points de surface du meilleur recalage de l'étalon. Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de position 3D des points de surface (en mm).

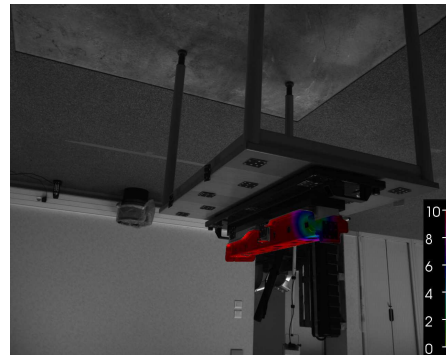
Reprojection du modèle



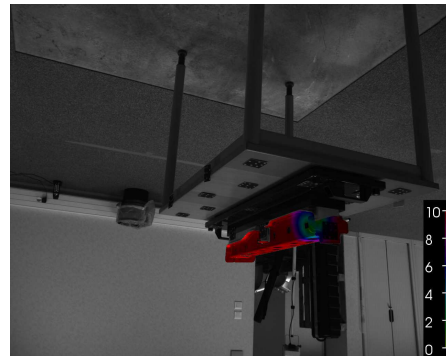
Erreur de reprojection



Méthode 2D/2D



Méthode 3D/2D

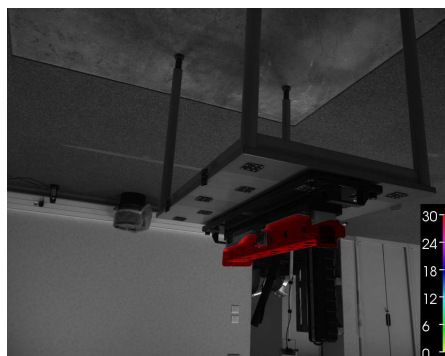


Méthode 3D/3D

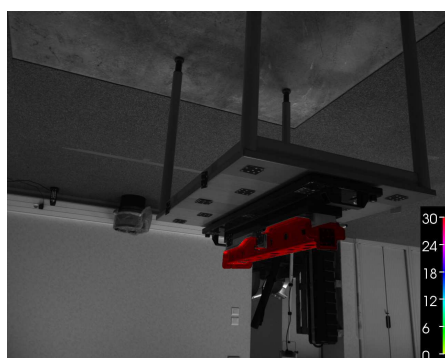
FIG. C.9 – Erreur de reprojection du plus mauvais recalage de l'étalon. *Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de reprojection associée (en pixels).*

Reprojection du modèle

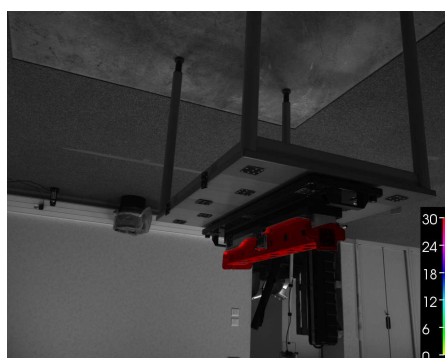
Erreur de position 3D



Méthode 2D/2D



Méthode 3D/2D

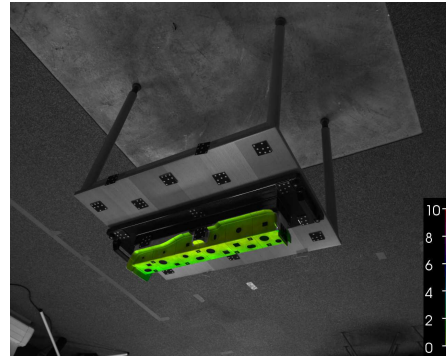


Méthode 3D/3D

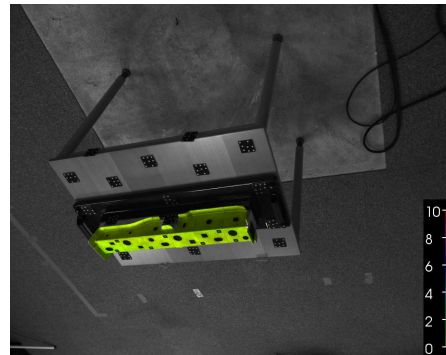
FIG. C.10 – Erreur de position 3D des points de surface du plus mauvais recalage de l'étalon. Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de position 3D des points de surface (en mm).

Reprojection du modèle

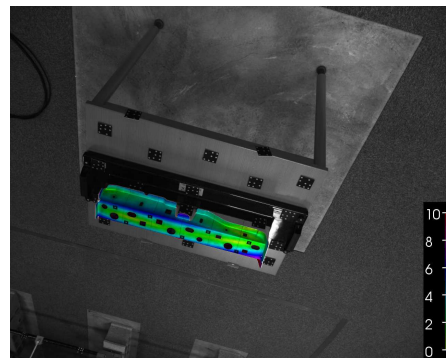
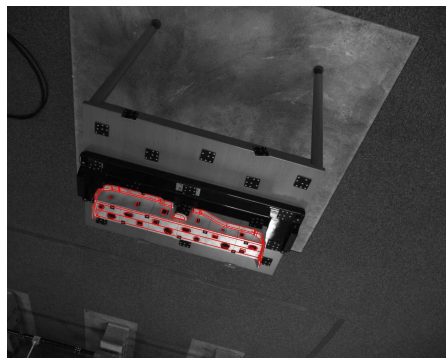
Erreur de reprojection



Méthode 2D/2D



Méthode 3D/2D

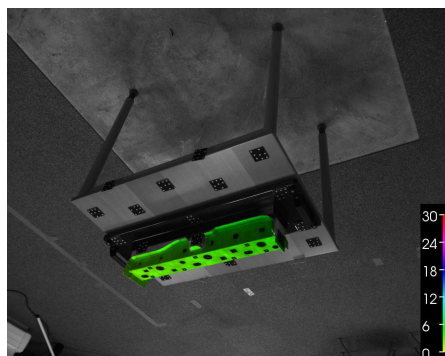


Méthode 3D/3D

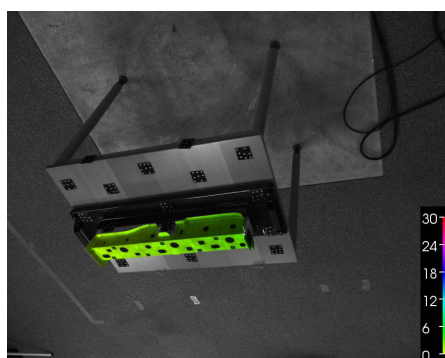
FIG. C.11 – Erreur de reprojection pour un recalage moyen de l'étalon. Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de reprojection associée (en pixels).

Reprojection du modèle

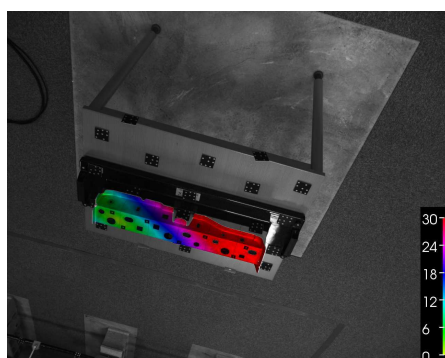
Erreur de position 3D



Méthode 2D/2D



Méthode 3D/2D



Méthode 3D/3D

FIG. C.12 – Erreur de position 3D des points de surface pour un recalage moyen de l'étalon. Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de position 3D des points de surface (en mm).



FIG. C.13 – Reprojection du modèle de l'étalon pour un recalage moyen selon la méthode 2D/2D.

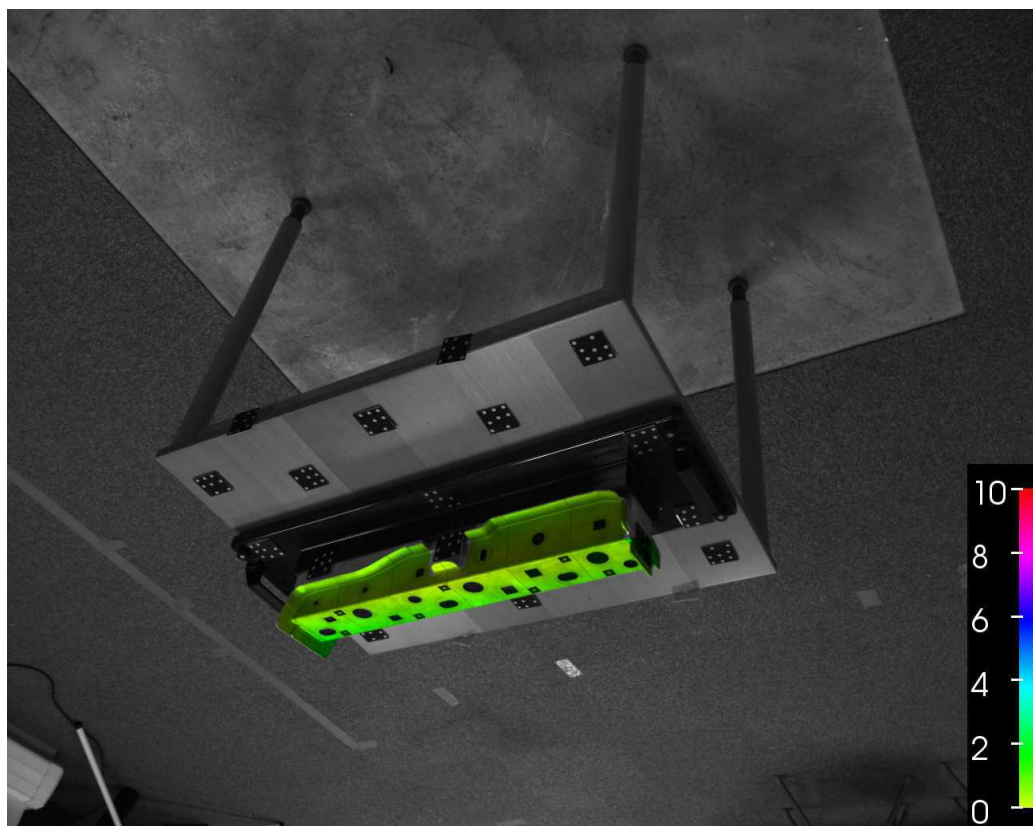


FIG. C.14 – Erreur de reprojection à la surface de l'étalon pour un recalage moyen selon la méthode 2D/2D. *L'échelle est donnée en pixels.*



FIG. C.15 – Erreur de position des points de surface de l'étalon pour un recalage moyen selon la méthode 2D/2D. *L'échelle est donnée en mm.*



FIG. C.16 – Reprojection du modèle de l'étalon pour un recalage moyen selon la méthode 3D/2D.

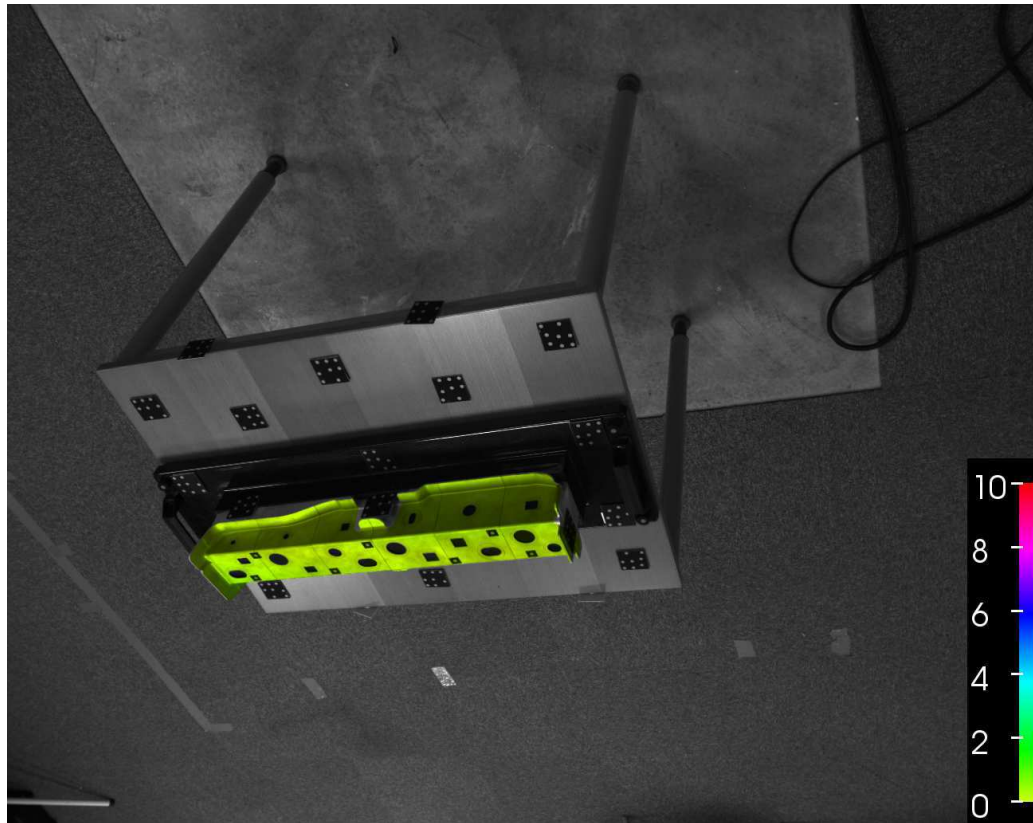


FIG. C.17 – Erreur de reprojection à la surface de l'étalon pour un recalage moyen selon la méthode 3D/2D. *L'échelle est donnée en pixels.*

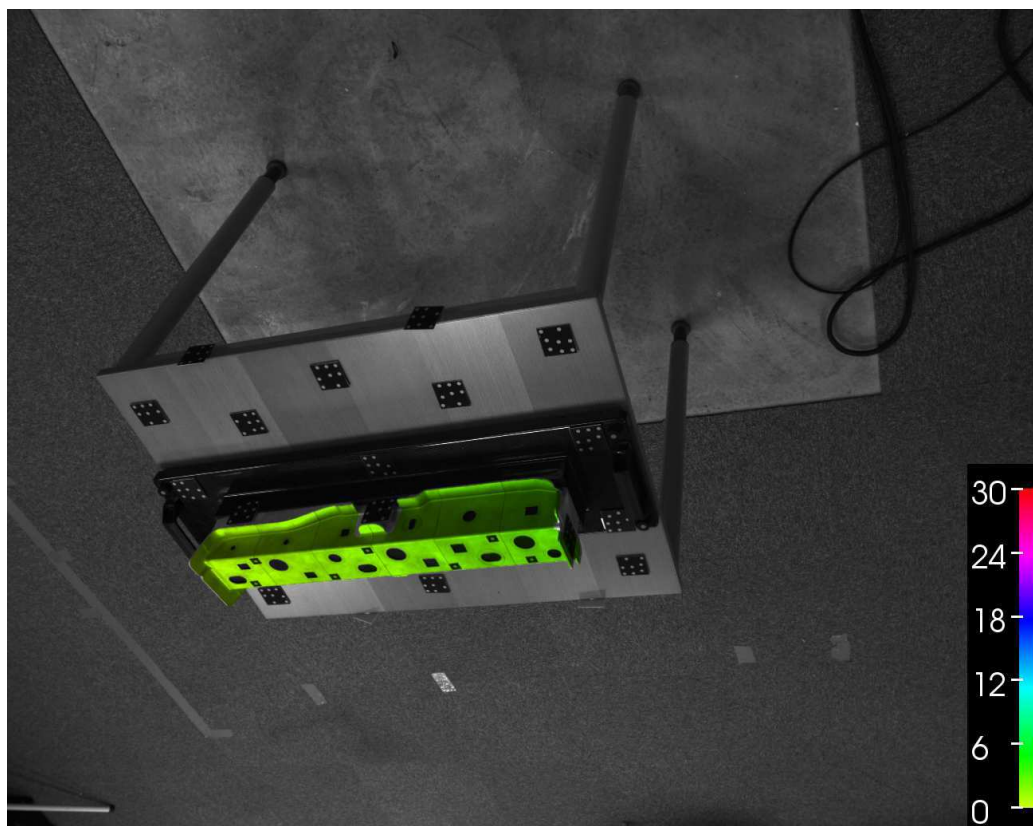


FIG. C.18 – Erreur de position des points de surface de l'étalon pour un recalage moyen selon la méthode 3D/2D. *L'échelle est donnée en mm.*

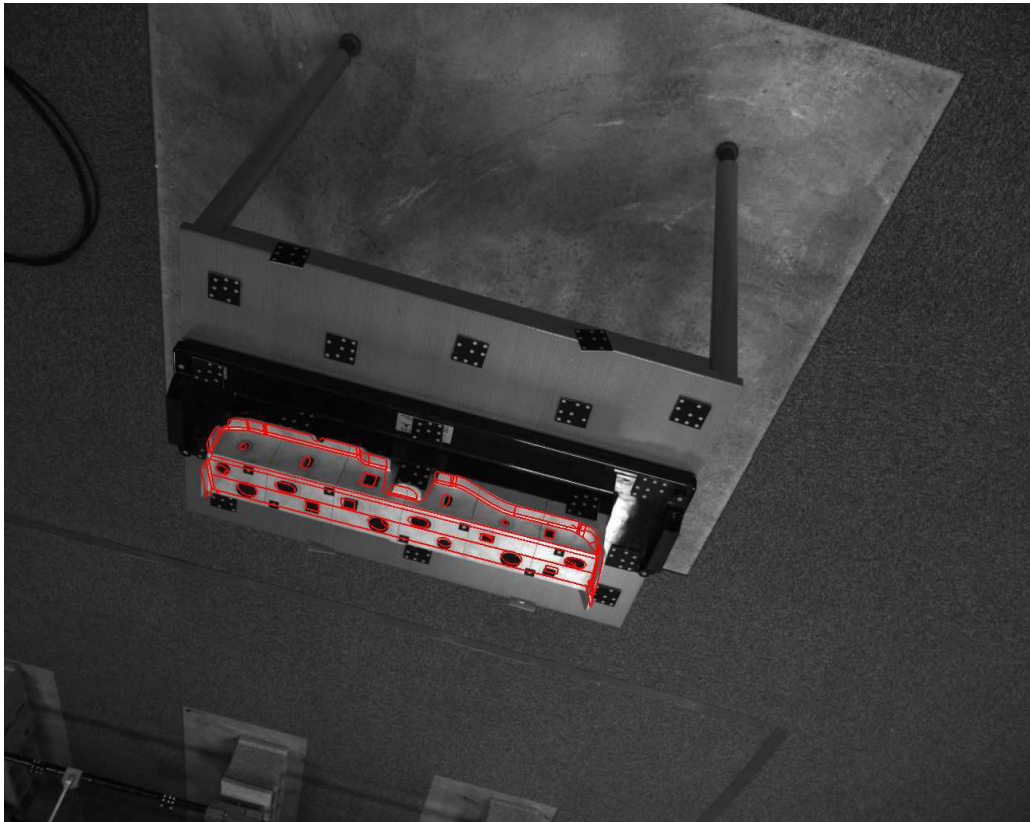


FIG. C.19 – Reprojection du modèle de l'étalon pour un recalage moyen selon la méthode 3D/3D.

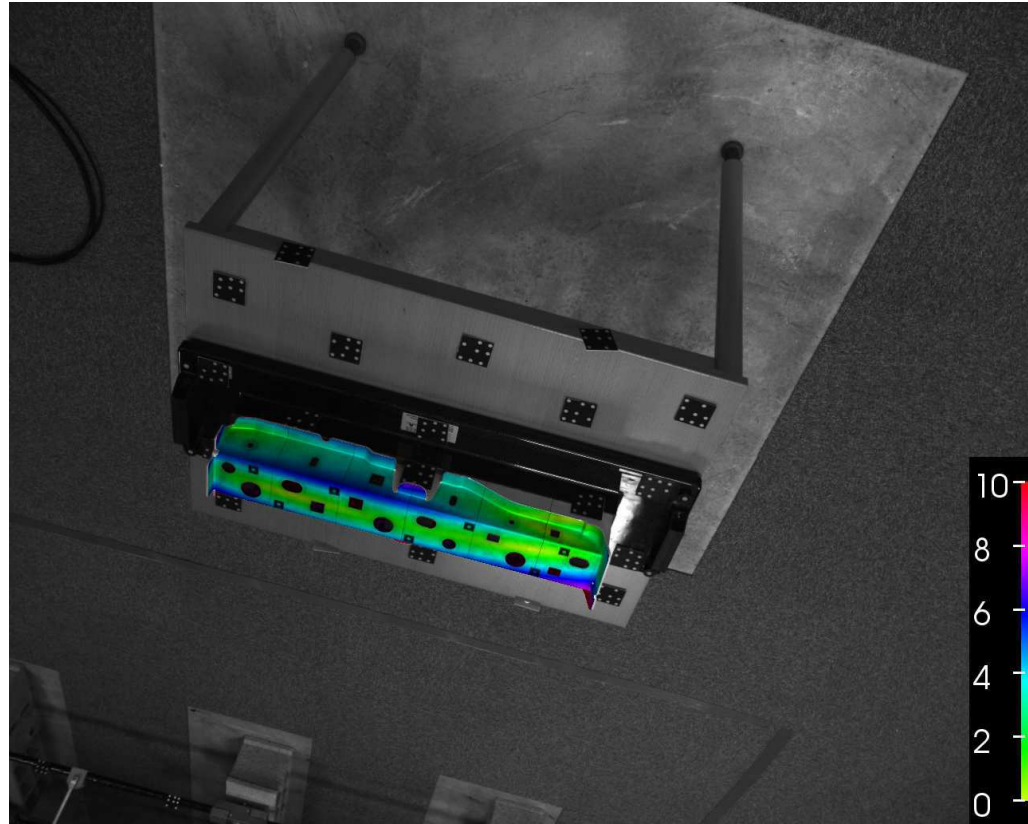


FIG. C.20 – Erreur de reprojection à la surface de l'étalon pour un recalage moyen selon la méthode 3D/3D. *L'échelle est donnée en pixels.*

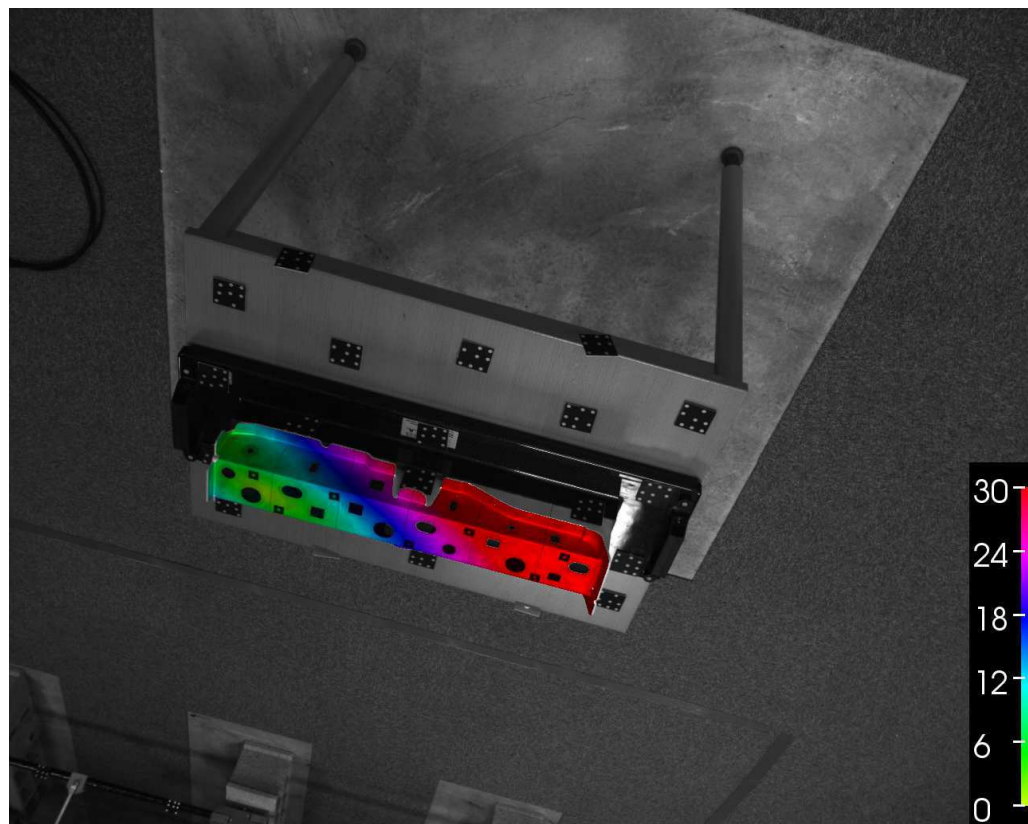


FIG. C.21 – Erreur de position des points de surface de l'étalon pour un recalage moyen selon la méthode 3D/3D. *L'échelle est donnée en mm.*

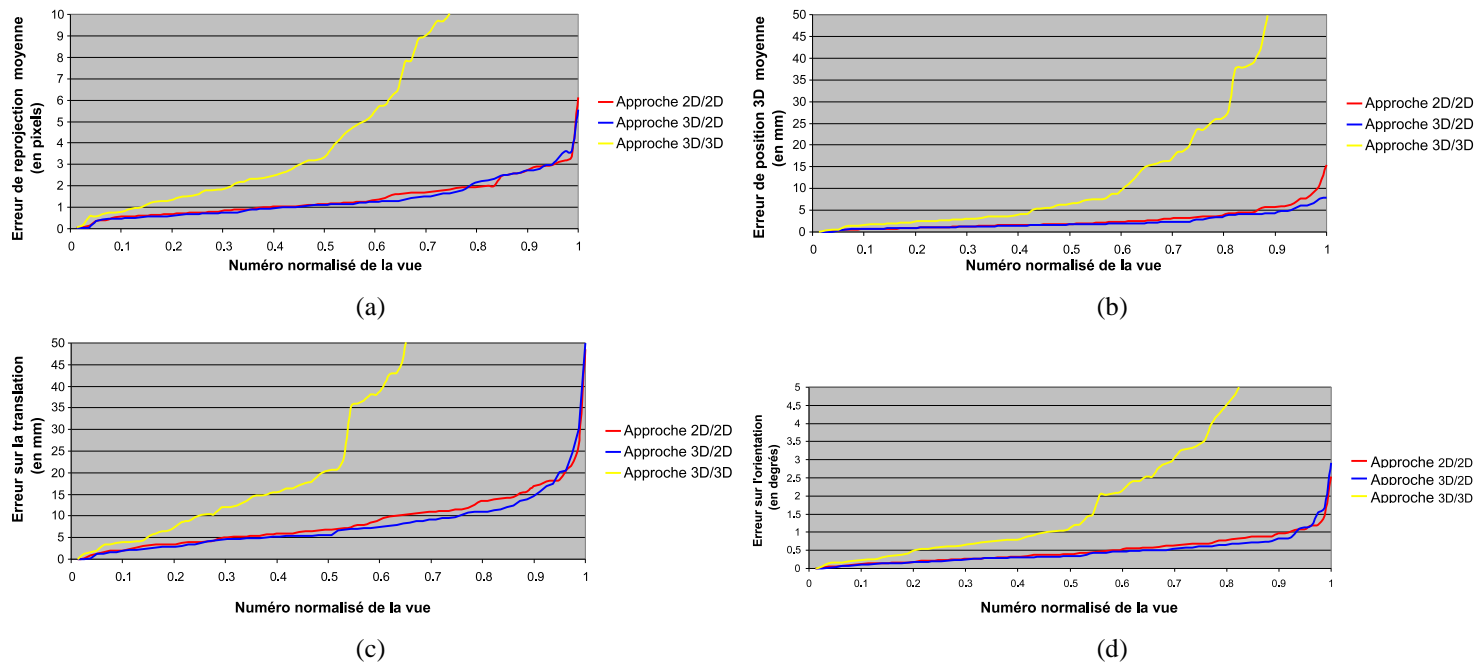


FIG. C.22 – Résultats de l'évaluation sur les 80 vues de la culasse. Le graphe (a) représente l'erreur de reprojection pour chacune des 80 vues, alors que le graphe (b) représente l'erreur de position 3D des points de surface, et les graphes (c) et (d) représentent respectivement l'erreur de translation et d'orientation de la pose. Pour chacun des graphes, les vues sont triées par ordre croissant d'erreur. L'abscisse des graphes correspond au numéro normalisé de la vue (ie la vue ayant l'erreur minimale correspond au numéro 0, la vue ayant l'erreur maximale correspond au numéro 1 et la vue ayant l'erreur médiane correspond à l'abscisse 0.5).

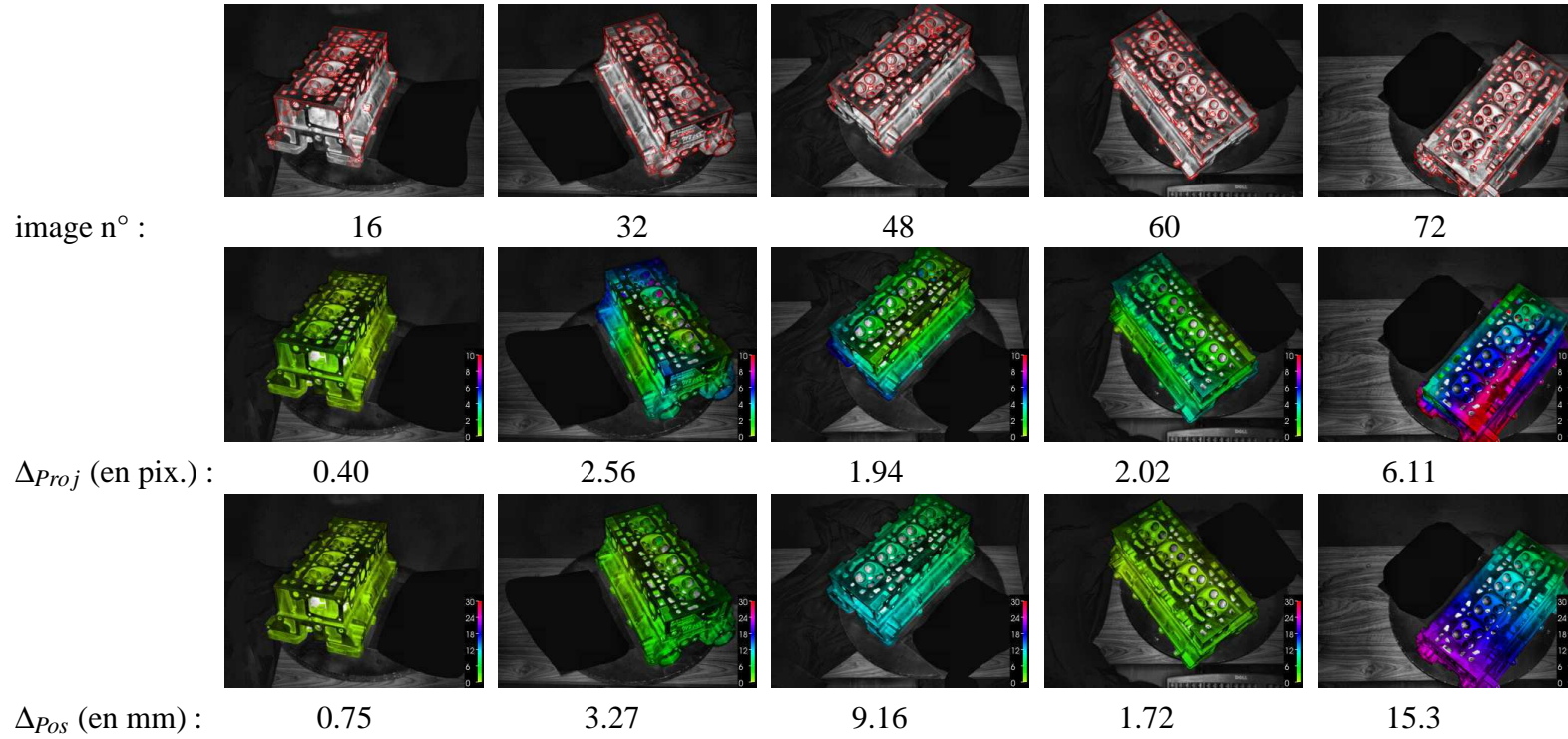


FIG. C.23 – Exemples de résultats pour la culasse avec la méthode 2D/2D. 1ère ligne d'images : reprojection de la culasse. 2ième ligne d'images : l'erreur de reprojection (en pixels) représentée en fausses couleurs. Δ_{proj} l'erreur de reprojection moyenne. 3ième ligne d'images : l'erreur de position 3D des points de surface (en mm) représentée en fausses couleurs. Δ_{pos} : l'erreur de position 3D moyenne.

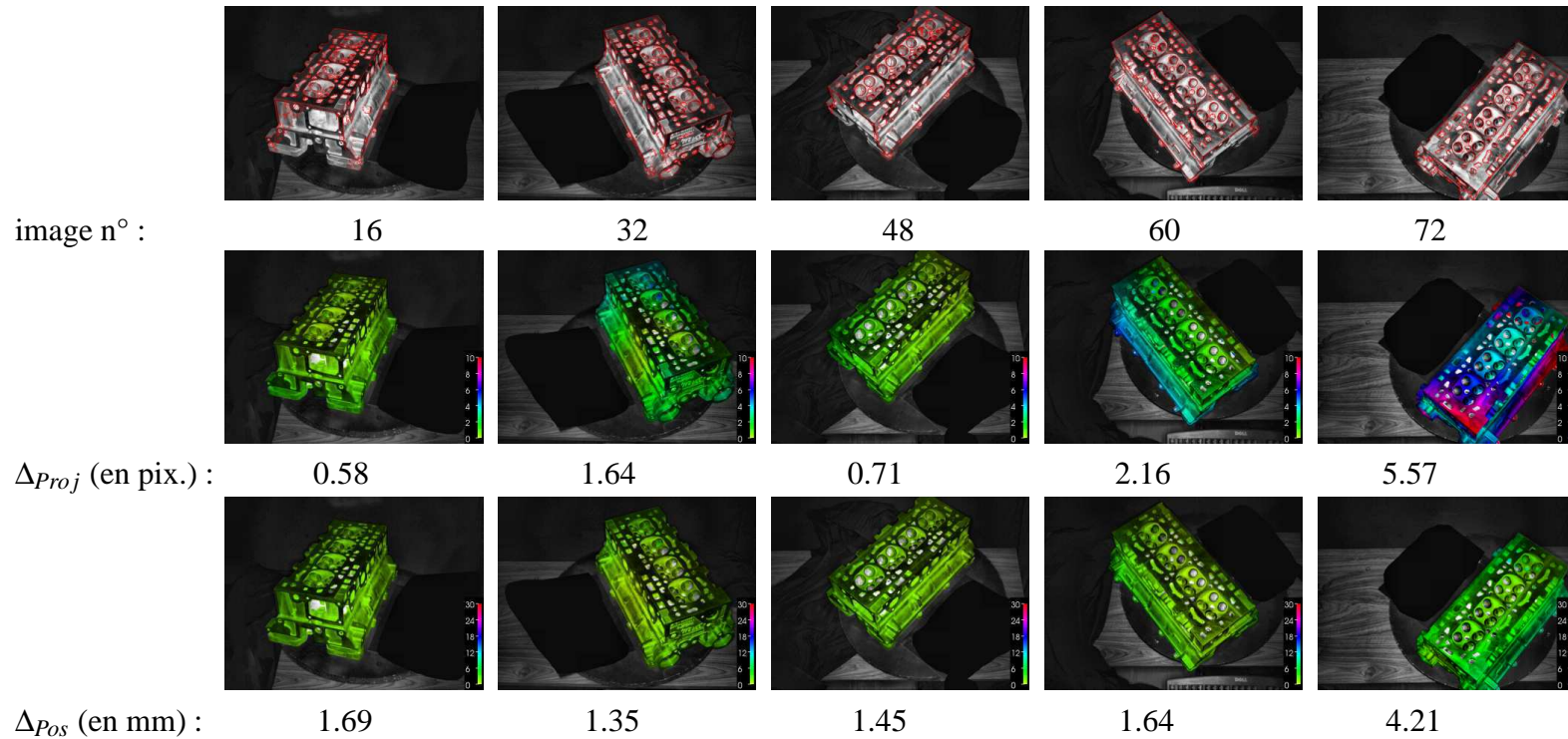


FIG. C.24 – Exemples de résultats pour la culasse avec la méthode 3D/2D. 1ère ligne d'images : reprojection de la culasse. 2ième ligne d'images : l'erreur de reprojection (en pixels) représentée en fausses couleurs. Δ_{proj} l'erreur de reprojection moyenne. 3ième ligne d'images : l'erreur de position 3D des points de surface (en mm) représentée en fausses couleurs. Δ_{pos} : l'erreur de position 3D moyenne.

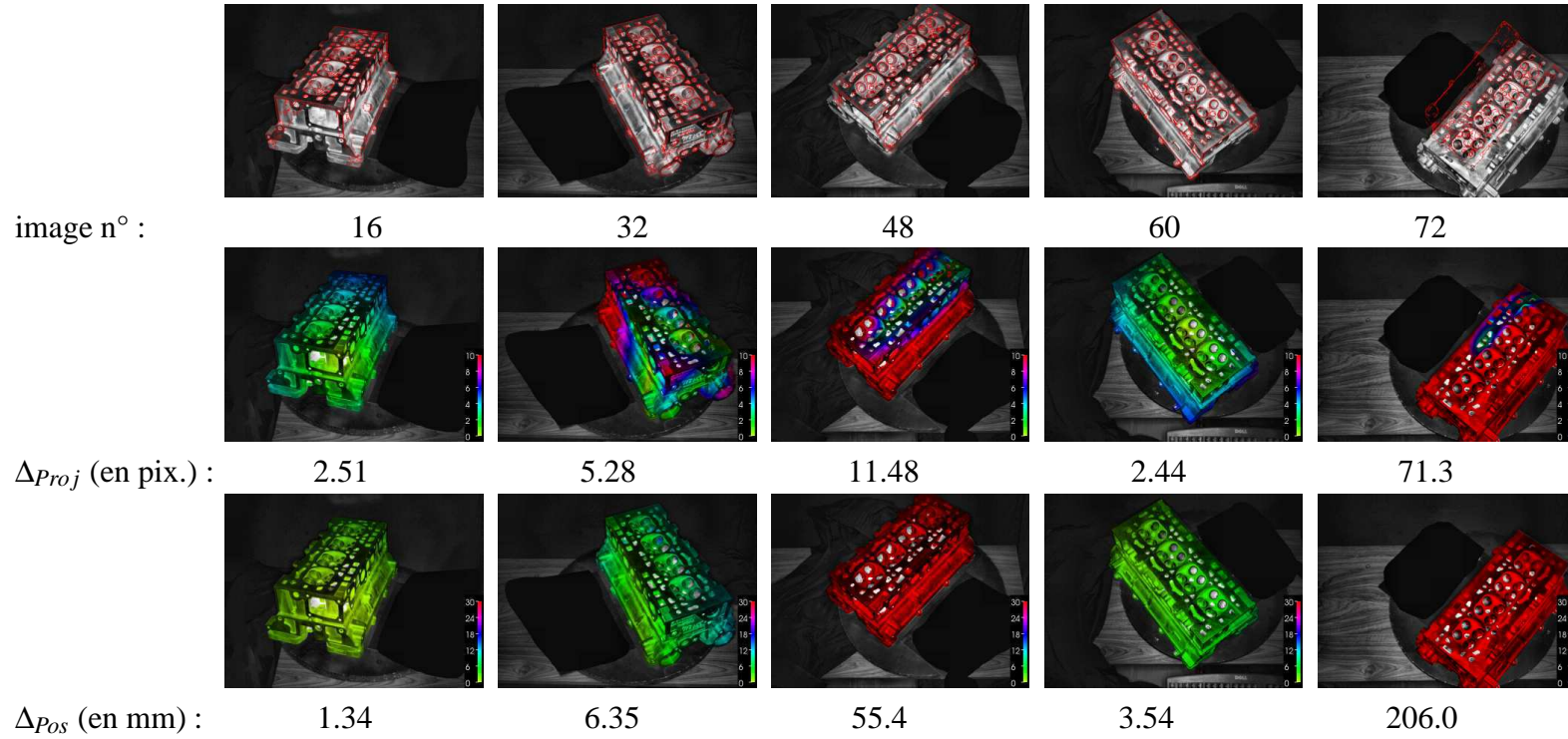


FIG. C.25 – Exemples de résultats pour la culasse avec la méthode 3D/3D. 1ère ligne d'images : reprojection de la culasse. 2ième ligne d'images : l'erreur de reprojection (en pixels) représentée en fausses couleurs. Δ_{proj} l'erreur de reprojection moyenne. 3ième ligne d'images : l'erreur de position 3D des points de surface (en mm) représentée en fausses couleurs. Δ_{pos} : l'erreur de position 3D moyenne.

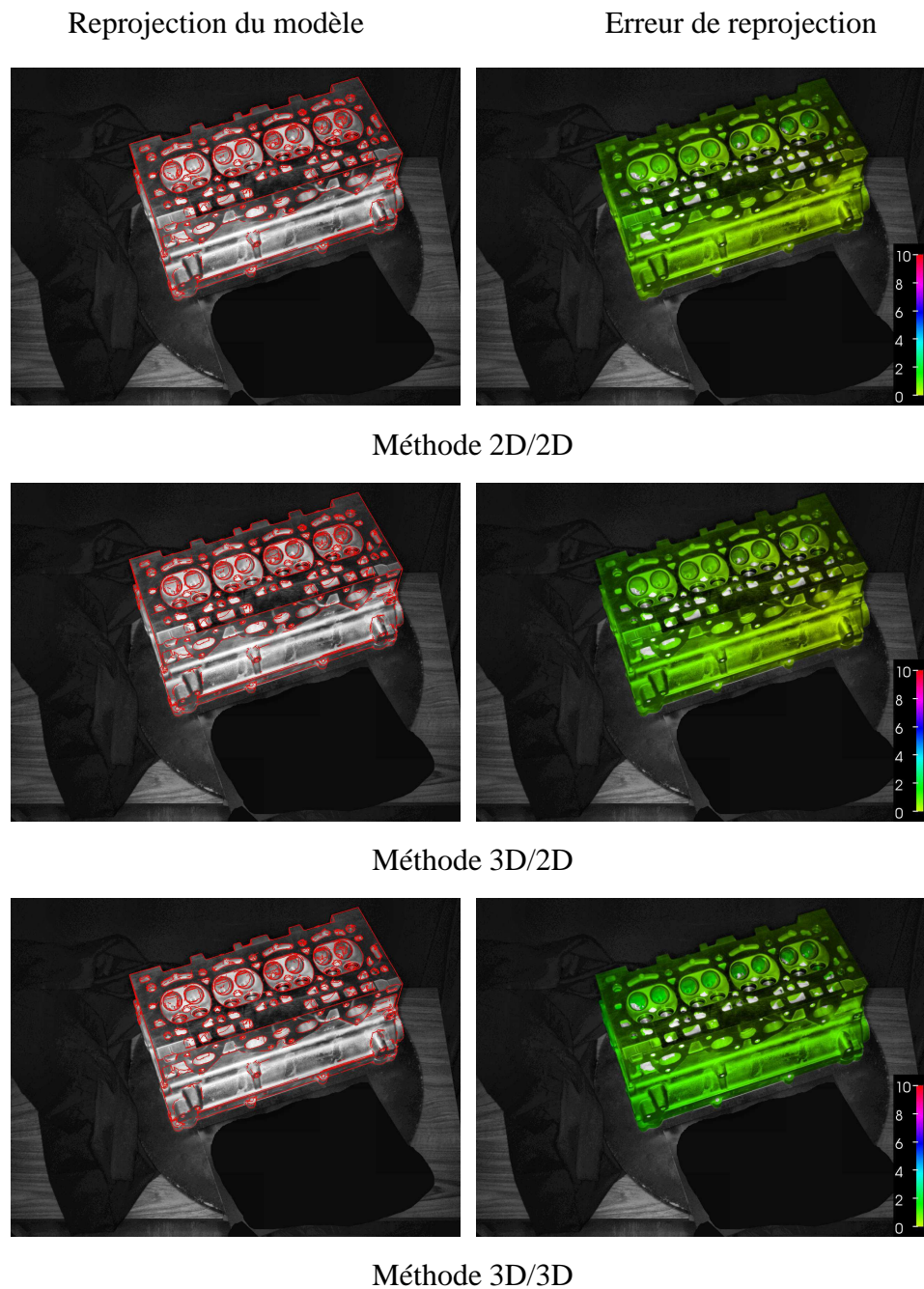
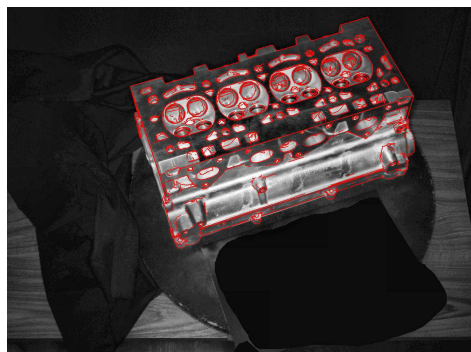
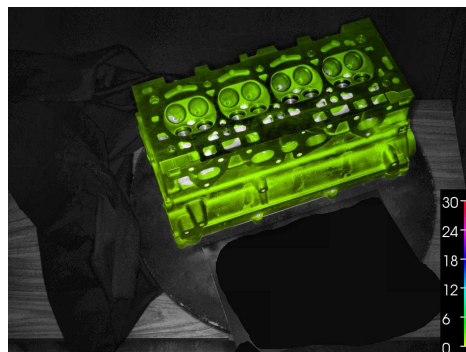


FIG. C.26 – Erreur de reprojection du meilleur recalage de la culasse. *Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de reprojection associée (en pixels).*

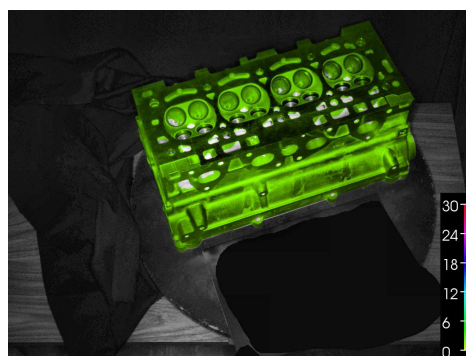
Reprojection du modèle



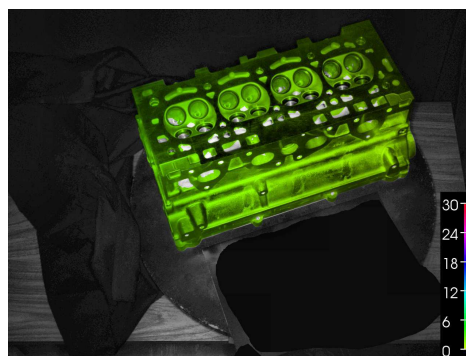
Erreur de position 3D



Méthode 2D/2D



Méthode 3D/2D

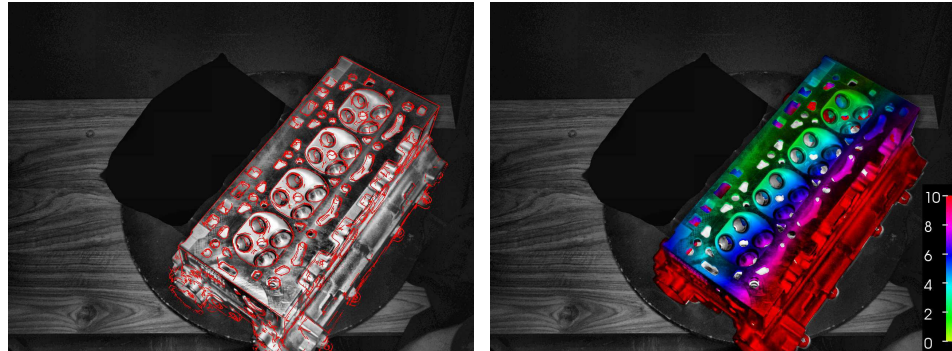


Méthode 3D/3D

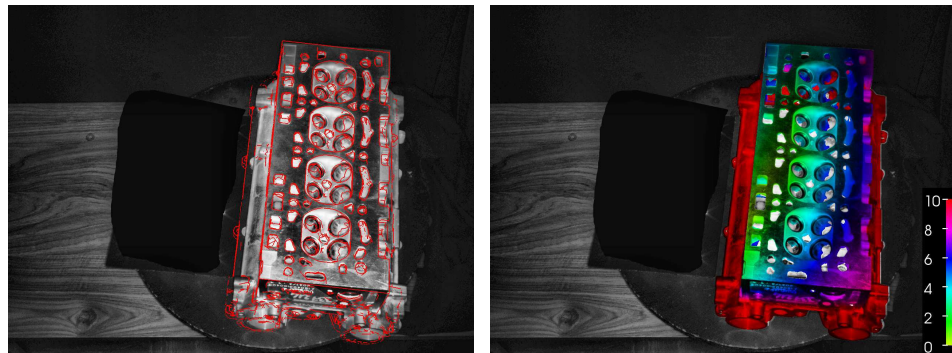
FIG. C.27 – Erreur de position 3D des points de surface du meilleur recalage de la culasse. *Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de position 3D des points de surface (en mm).*

Reprojection du modèle

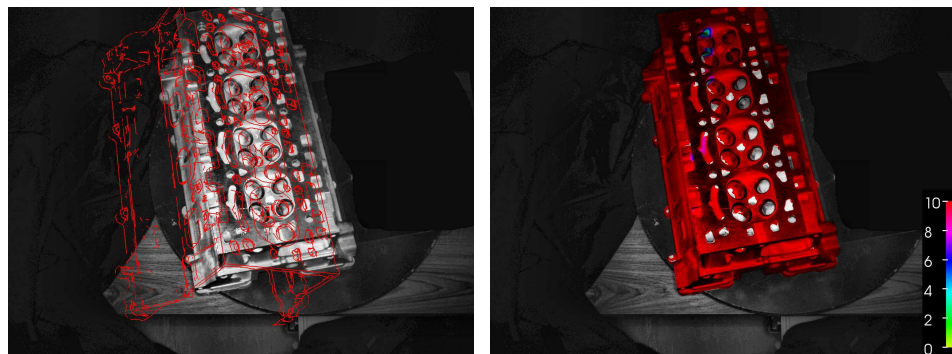
Erreur de reprojection



Méthode 2D/2D



Méthode 3D/2D

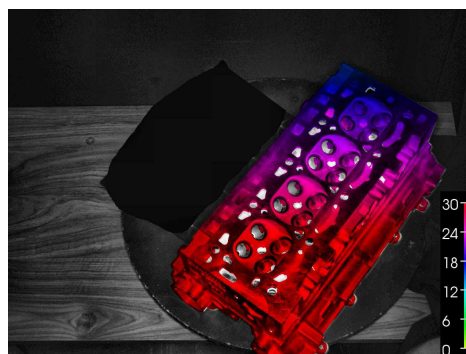
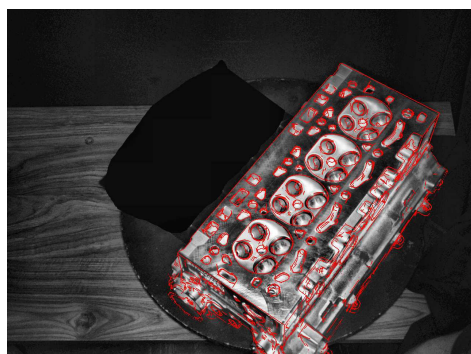


Méthode 3D/3D

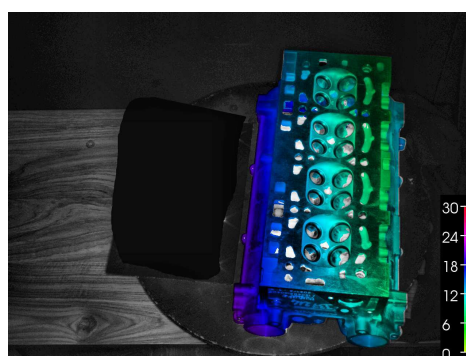
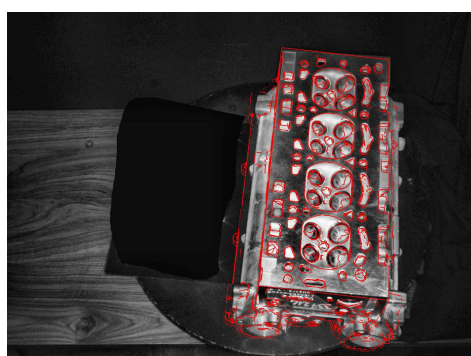
FIG. C.28 – Erreur de reprojection du plus mauvais recalage de la culasse. *Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de reprojection associée (en pixels).*

Reprojection du modèle

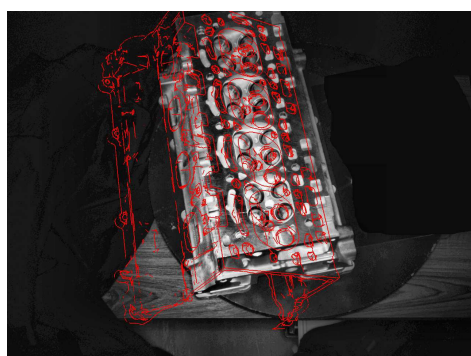
Erreur de position 3D



Méthode 2D/2D



Méthode 3D/2D

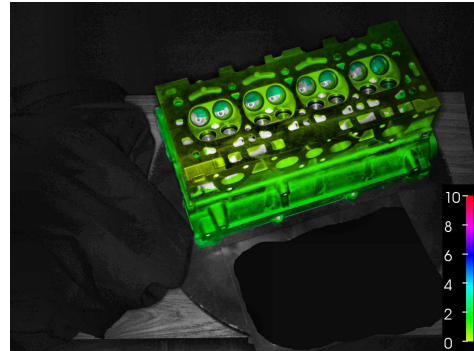
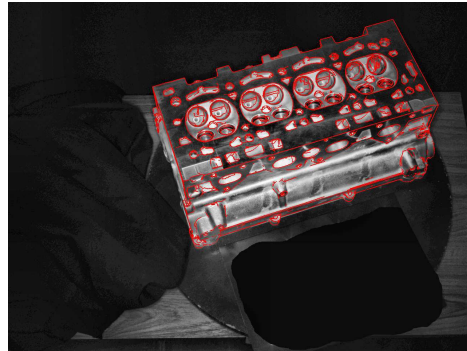


Méthode 3D/3D

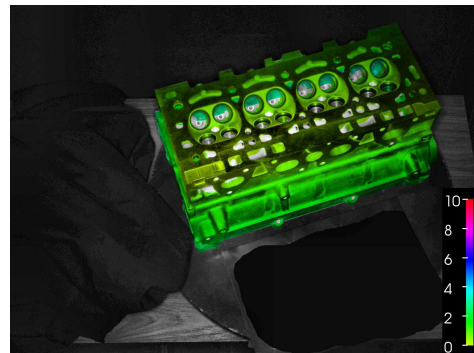
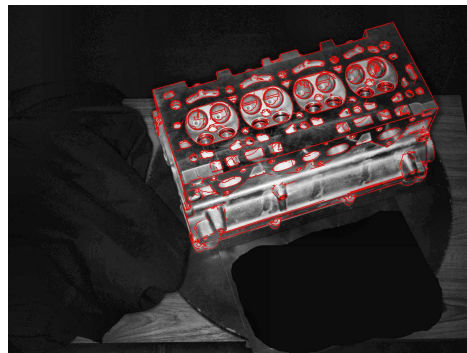
FIG. C.29 – Erreur de position 3D des points de surface du plus mauvais recalage de la culasse. Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de position 3D des points de surface (en mm).

Reprojection du modèle

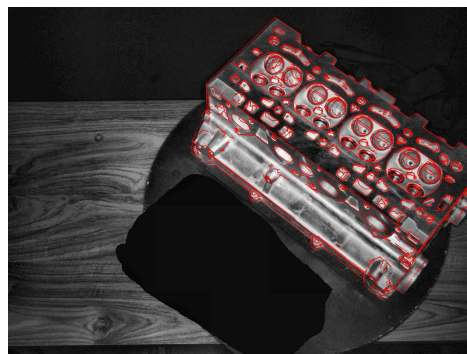
Erreur de reprojection



Méthode 2D/2D



Méthode 3D/2D

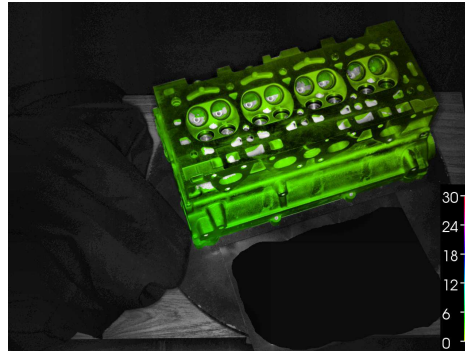
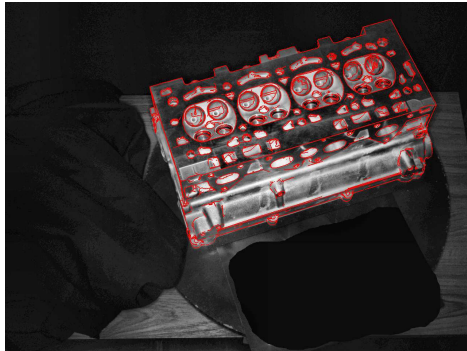


Méthode 3D/3D

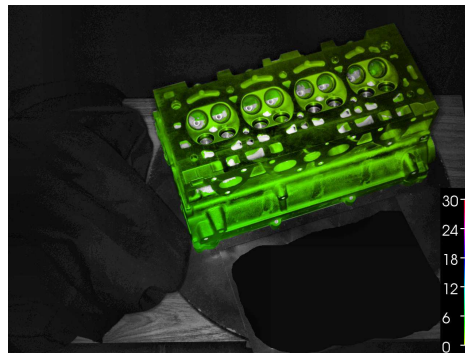
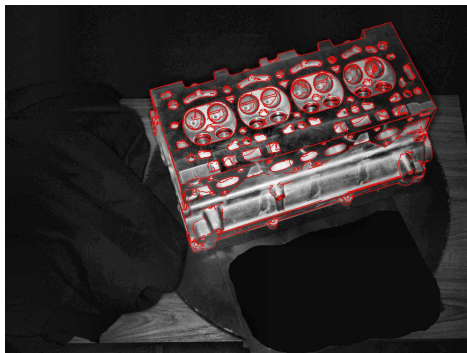
FIG. C.30 – Erreur de reprojection pour un recalage moyen de la culasse. *Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de reprojection associée (en pixels).*

Reprojection du modèle

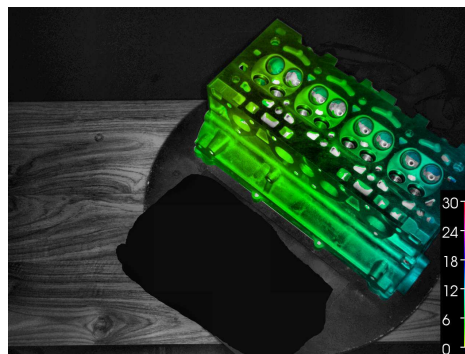
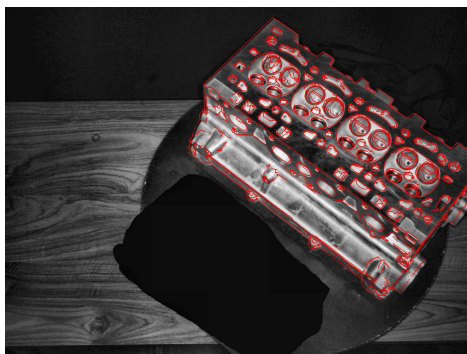
Erreur de position 3D



Méthode 2D/2D



Méthode 3D/2D



Méthode 3D/3D

FIG. C.31 – Erreur de position 3D des points de surface pour un recalage moyen de la culasse. Chaque ligne d'images correspond à une méthode différente. L'image de gauche correspond à la reprojection du modèle CAO et l'image de droite représente l'erreur de position 3D des points de surface (en mm).

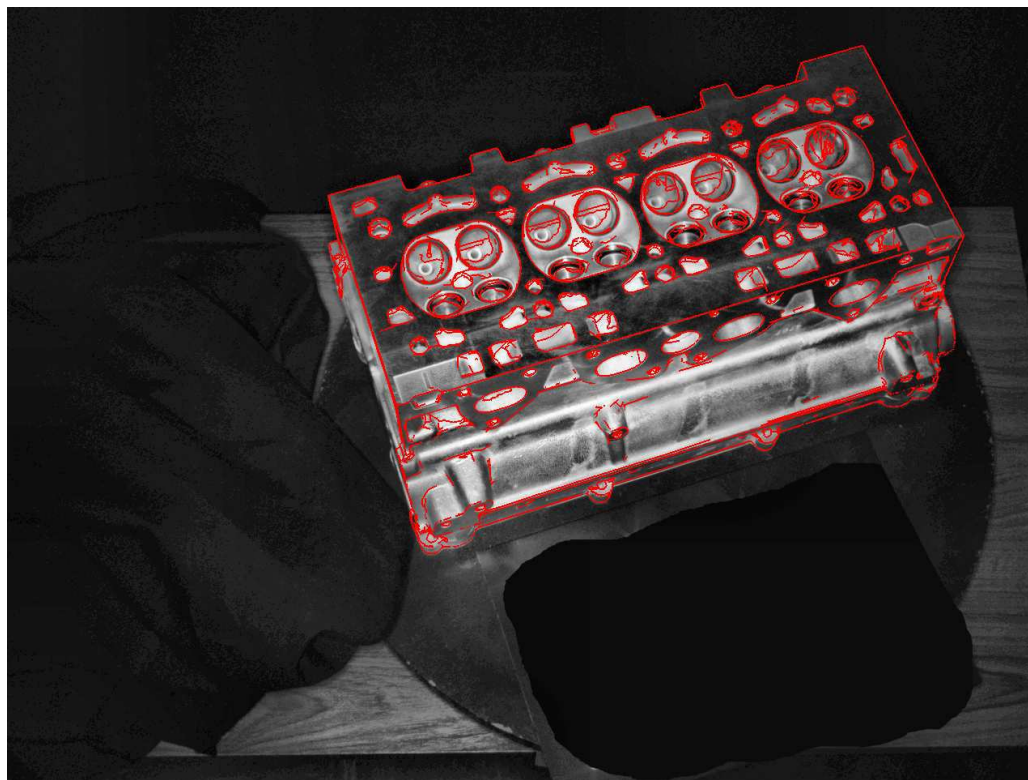


FIG. C.32 – Reprojection du modèle de la culasse pour un recalage moyen selon la méthode 2D/2D.

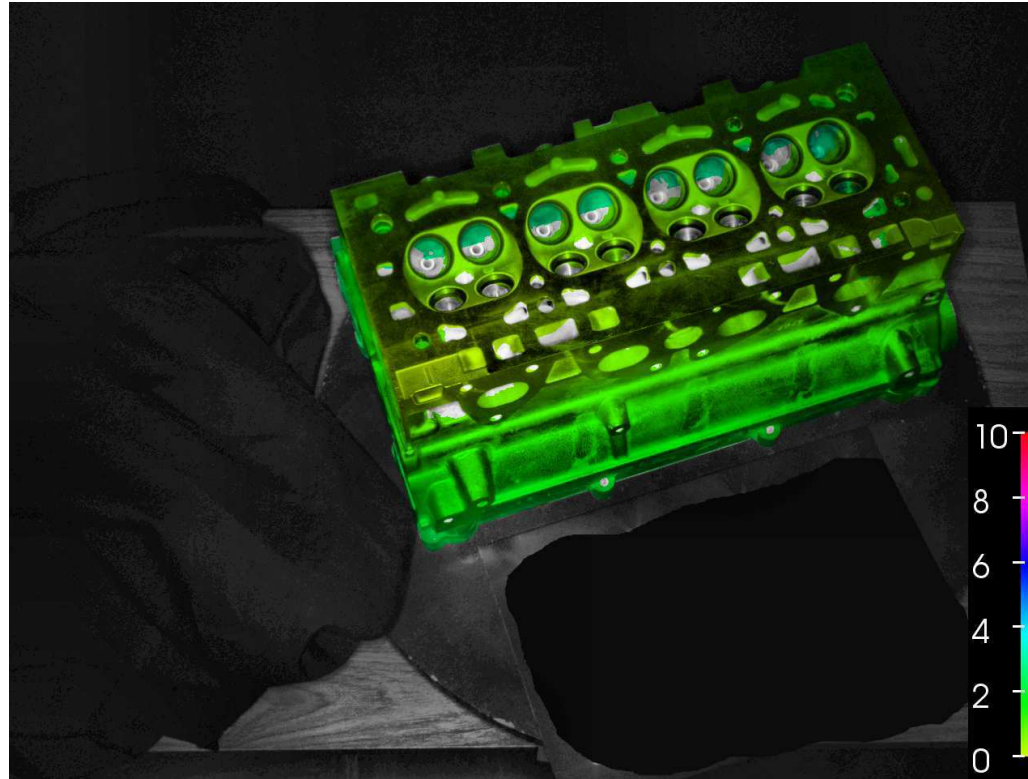


FIG. C.33 – Erreur de reprojection à la surface de la culasse pour un recalage moyen selon la méthode 2D/2D. *L'échelle est donnée en pixels.*

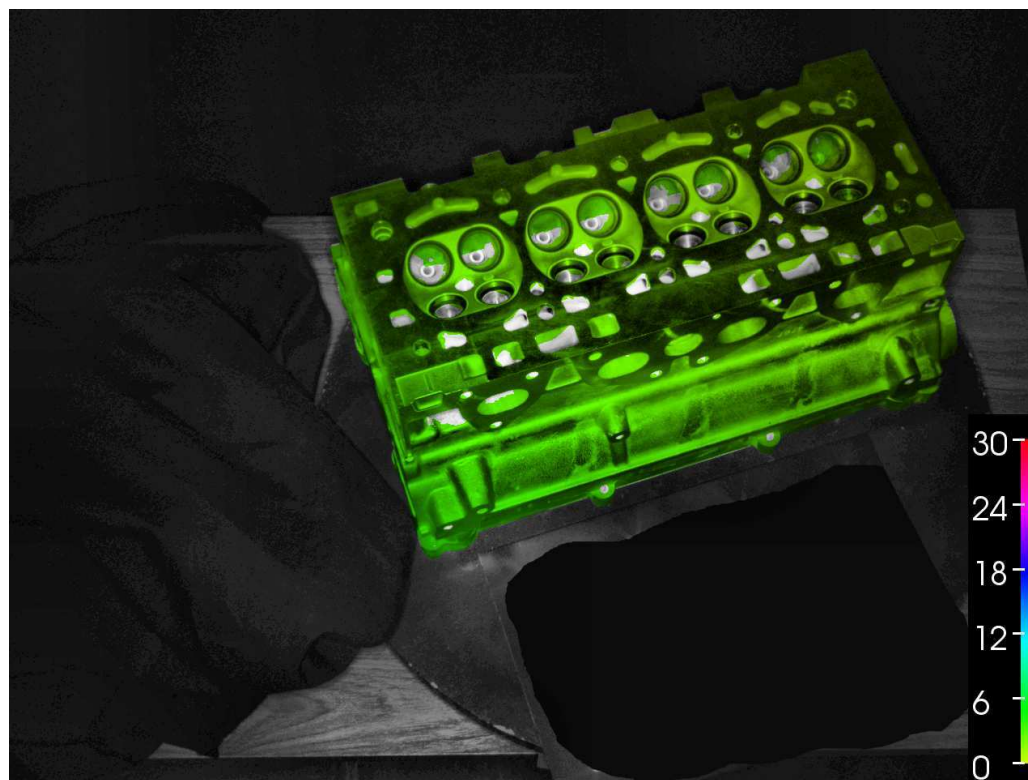


FIG. C.34 – Erreur de position des points de surface de la culasse pour un recalage moyen selon la méthode 2D/2D. *L'échelle est donnée en mm.*



FIG. C.35 – Reprojection du modèle de la culasse pour un recalage moyen selon la méthode 3D/2D.

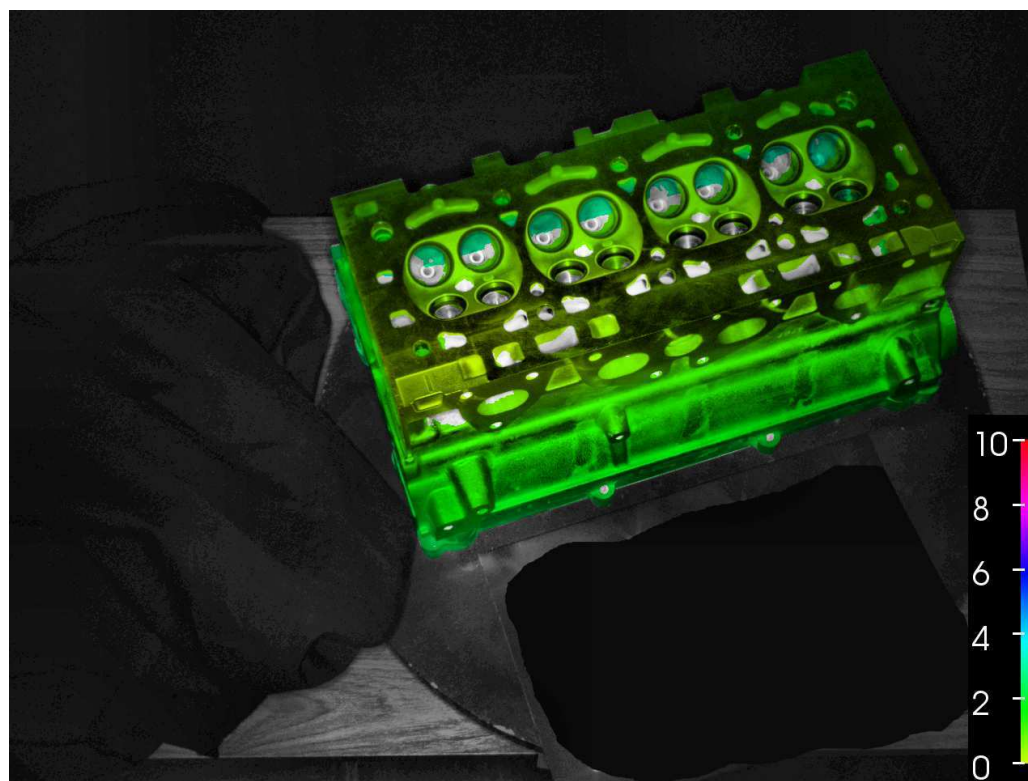


FIG. C.36 – Erreur de reprojection à la surface de la culasse pour un recalage moyen selon la méthode 3D/2D. *L'échelle est donnée en pixels.*



FIG. C.37 – Erreur de position des points de surface de la culasse pour un recalage moyen selon la méthode 3D/2D. *L'échelle est donnée en mm.*

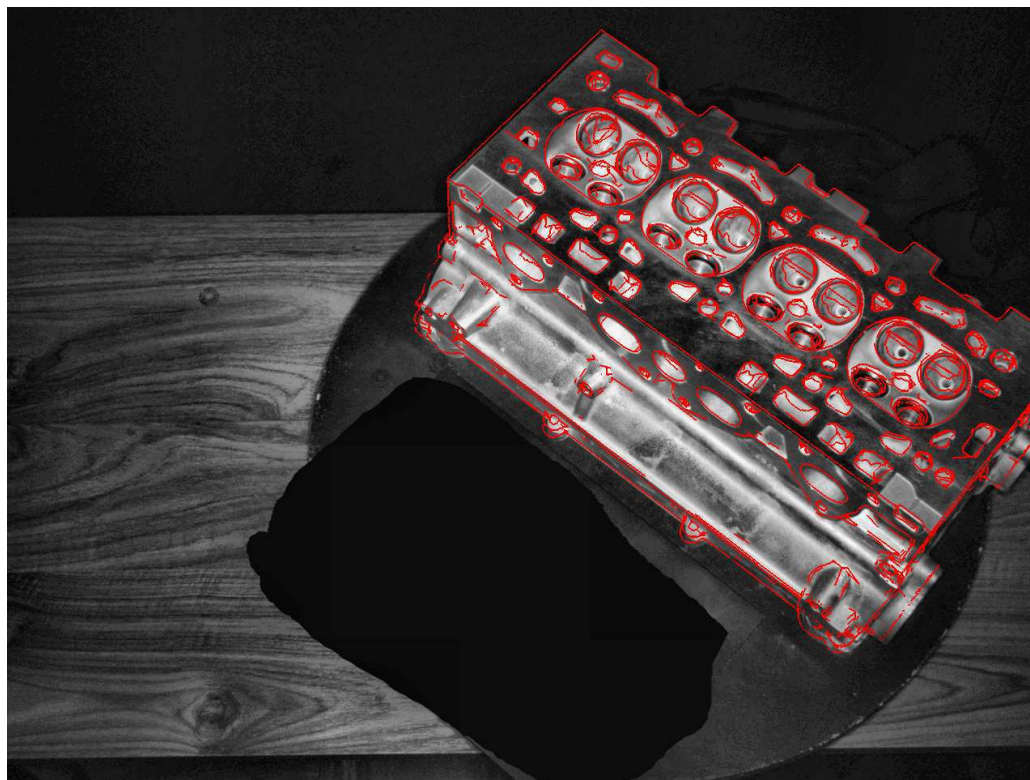


FIG. C.38 – Reprojection du modèle de la culasse pour un recalage moyen selon la méthode 3D/3D.

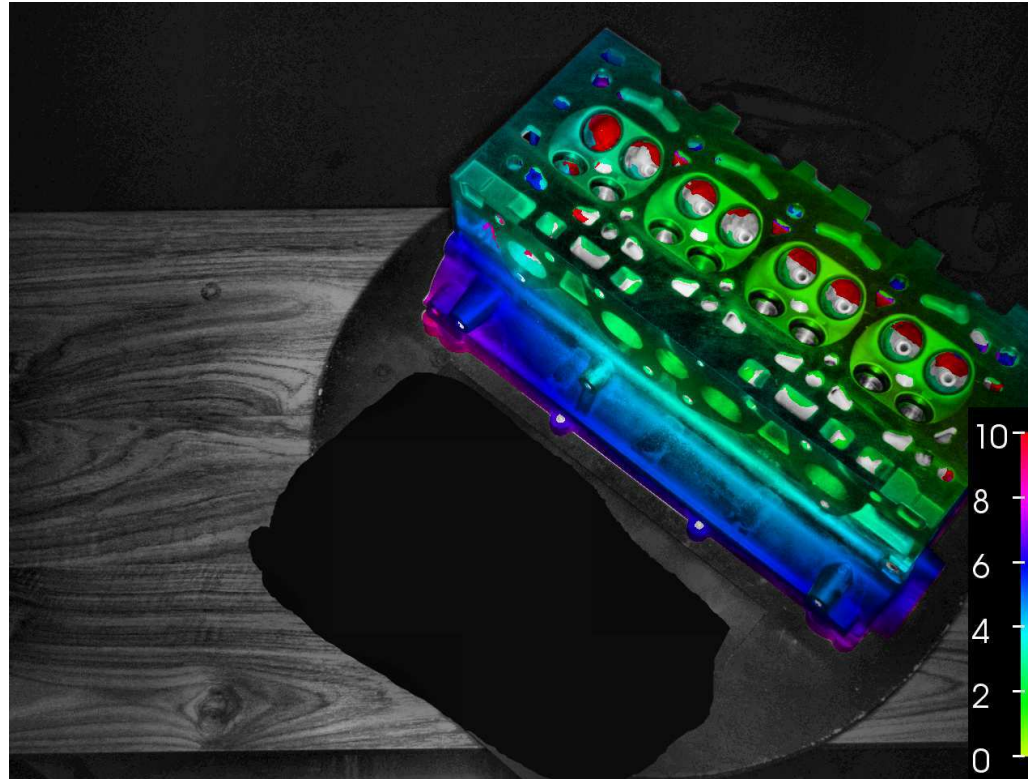


FIG. C.39 – Erreur de reprojection à la surface de la culasse pour un recalage moyen selon la méthode 3D/3D. *L'échelle est donnée en pixels.*



FIG. C.40 – Erreur de position des points de surface de la culasse pour un recalage moyen selon la méthode 3D/3D. *L'échelle est donnée en mm.*

Alignement d'Objets Mécaniques Complexes par Vision Monoculaire

Ce mémoire présente un système de recalage visuel adapté aux pièces mécaniques. Il s'agit d'estimer, de manière automatique et précise, la position et l'orientation 3D d'un objet connu, à partir de son image vidéo. Les objets considérés possèdent la particularité d'être complexes, généralement métalliques et non-texturés.

La solution proposée exploite un modèle d'objet caractérisant à la fois l'apparence et la géométrie de l'objet. Afin de pouvoir estimer la pose de l'objet, nous proposons un processus de mise en correspondance exploitant conjointement les informations d'apparence et de géométrie du modèle. Cette étape exploite la capacité à estimer une pose approximative de l'objet à partir d'un seul appariement pour filtrer efficacement les faux appariements. Un processus d'estimation précise de la pose est ensuite proposé.

L'évaluation du processus complet sur des pièces mécaniques issues de l'industrie souligne la précision et la robustesse de ce dernier.

Registration of complex mechanic objects by monocular vision

This thesis presents a visual registration system suitable for mechanical objects. The aim of this work is to perform an accurate and automatic estimation of the 3D position and orientation of a known 3D object, from a single camera image. The studied objects are complex, generally metallic and textureless.

To solve this problem, we introduce an object model which characterizes both the visual aspect and the 3D geometry of the object. To estimate the object position and orientation, we propose a matching process which exploits both visual and geometric data of the object model. Because model features allow computing a coarse pose of the object from a single match, false matches can be filtered efficiently. A pose estimation process is then explained.

Finally, an experimental evaluation of the complete visual registration process is exposed. This evaluation was led on industrial mechanic objects (a car door, an automotive cylinder-head...). The results underline the accuracy and the robustness of the proposed solution.